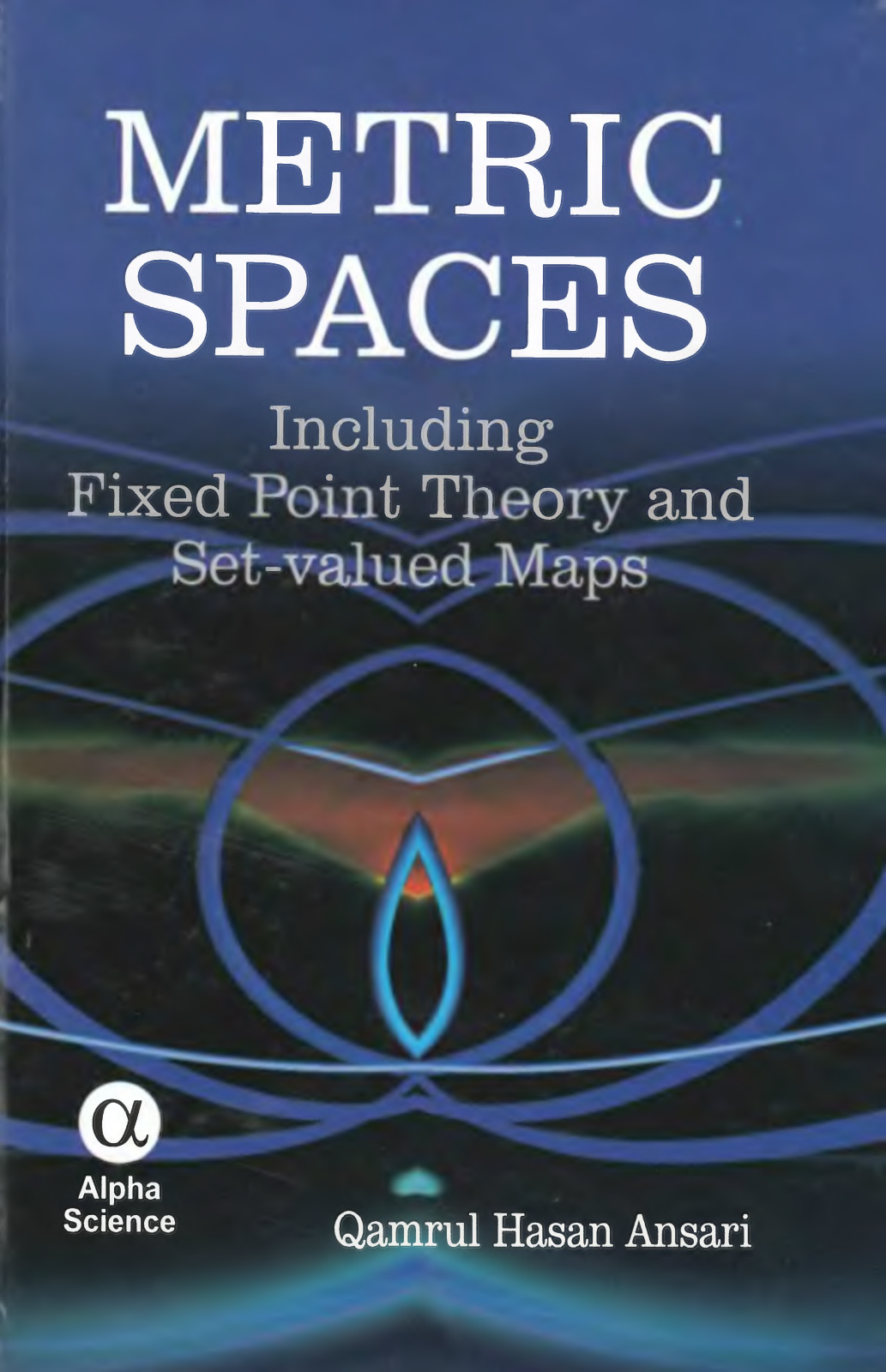
METRIC SPACES

Including Fixed Point Theory and Set-valued Maps



Alpha Science

Qamrul Hasan Ansari

METRIC SPACES

Including
Fixed Point Theory and
Set-valued Maps

METRIC SPACES

Including Fixed Point Theory and Set-valued Maps

Qamrul Hasan Ansari

Alpha Science International Ltd.
Oxford, U.K.

METRIC SPACES
Including Fixed Point Theory and Set-valued Maps
208 pgs.

Qamrul Hasan Anasari
Department of Mathematics
Aligarh Muslim University
Aligarh

Copyright © 2010

ALPHA SCIENCE INTERNATIONAL LTD.
7200 The Quorum, Oxford Business Park North
Garsington Road, Oxford OX4 2JZ, U.K.

www.alphasci.com

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior written permission of the publisher.

Printed from the camera-ready copy provided by the Author.

ISBN 978-1-84265-655-6

Printed in India

Dedicated to my wife Shabista Begum
and
our daughters Reem, Warda and Suzan

This page intentionally left blank

Preface

This book is based on a course on Metric Spaces that I have been teaching to the third year B.Sc. (Honours) Mathematics students at the Aligarh Muslim University, Aligarh, for several years. It is the essence of my lecture notes on the subject as well as of the several lectures on set-valued maps, Ekeland's variational principle and equilibrium problems that I have delivered in several workshops and symposia. The course on metric spaces is the first course leading to the study of functional analysis. It provides the idea of a distance between any two points on an arbitrary set by generalizing three basic properties of a distance between two points. These are: (i) the distance between any two points is always positive, and it vanishes if and only if two points lie on the same position, (ii) the distance between two points is the same from whichever point it is measured, and (iii) the distance from one point to another cannot exceed the sum of the distances from these points to an intermediate point. Numerous books on metric spaces and topology containing metric spaces are available but in this book, many definitions and the proof of the results are furnished by providing figures. More than half of the exercises are supported by the solutions or hints. No existing book on metric spaces has set-valued maps, Ekeland's variational principle, Caristi's fixed point theorem, Takahashi's minimization theorem and equilibrium problems. However, this is the first book in which the equilibrium problem, a unified model of several problems, namely, optimization problems, saddle point problems, fixed point problems, complementarity problems, variational inequality problems, Nash equilibrium problems, etc, is presented in the setting of metric spaces. An elementary treatment of the set-valued maps is given. The Banach contraction theorem for set-valued maps (known as Nadler's fixed point theorem), Caristi's fixed point theorem and several related results for set-valued maps are included. Though the originality of the proofs is not guaranteed, but the simplicity and clarity of the proofs is definitely ensured.

In Chapter 1, we present the basic definitions, examples and results concerning metric spaces. The concepts of open spheres, closed spheres and metric subspaces are also given in this chapter. In Chapter 2, convergence of the sequences, Cauchy sequences and the completion of metric spaces are discussed. One of the useful results from functional analysis is Cantor's intersection section theorem which is also presented in this chapter. Some topological concepts, namely, countability, separability and Baire's category theorem are given in Chapter 3. Chapter 4 deals with the concepts of compact spaces, sequentially compact spaces and Bolzano-Weierstrass property. The equivalence of these three concepts is also discussed. In Chapter 5, continuity of functions is defined on a metric space to another metric space. Along with several characterizations of continuity, the image of a compact set under continuous function is also discussed. The concepts of homeomorphism and equivalent metrics are included. The uniform convergence of sequences of functions is given.

Chapter 6 deals with the connected metric spaces, components and their properties. The continuous image of a connected set is also discussed and as a consequence, the intermediate theorem and a fixed point theorem are derived. Chapter 7 is devoted to the metric fixed point theory. Banach contraction theorem and several of its generalizations along with their applications are presented. Caristi's fixed point theorem and some of its consequences are discussed. An introduction of set-valued analysis, continuity and fixed point theory of set-valued maps are given in Chapter 8. One of the most useful and important results from nonlinear analysis is Ekeland's variational principle. This principle along with several of its equivalent forms and Takahashi's minimization theorem are given in Chapter 9. Here, an introduction of the theory of equilibrium problems is presented. The equilibrium version of Ekeland's variational principle and several of its equivalent forms are also presented in this last chapter. The book is intended to be of use to academics and undergraduate students.

I would like to express my profound thanks and gratitude to my friends Prof. S. Al-Homidan, Saudi Arabia, Prof. M. A. Khamsi, U.S.A. and Prof. J.-C. Yao, Taiwan, for their kind encouragement at different times during the preparation of this book.

I would also like to express my sincere thanks to my esteemed colleagues Prof. M. Imdad, Dr. Rais Ahmad, Dr. Shamshad Husain and Dr. S. C. Gupta and our students Mr. M. Dilshad and Miss Himanshu Gupta for their careful reading of the manuscript. Also, it is my pleasant duty to thank Mr. N. K. Mehra, Publisher and Managing Director, Narosa Publishing House Pvt. Ltd., New Delhi, for taking keen interest in getting this book published for the benefit of students and workers in the field.

Last, but not the least, I wish to thank my family for their infinite patience, encouragement and forbearance. This work would not have seen the light of day had it not been for their unflinching support during this project. I remain indebted to them.

I welcome any opinions, suggestions, and added information which will improve the future editions and help readers in future. Benefits for readers will be the best reward for the author.

Qamrul Hasan Ansari

Notations and Abbreviations

$\mathbb{K}$	the set of all real or complex numbers
$\mathbb{R}$	the set of all real numbers
$\mathbb{C}$	the set of all complex numbers
$\mathbb{N}$	the set of all natural numbers
$\mathbb{Z}$	the set of all integers
$\mathbb{Q}$	the set of all rational numvers
$\overline{\mathbb{R}}$	the extended real line
$\mathbb{R}^n$	the n-dimensional Euclidean space
$\mathbb{C}^n$	the space of ordered n-tuples complex numbers
$\mathbb{R}_+$	$[0,\infty)$
ℓ^∞	the space of all bounded sequences of real or complex numbers
ℓ^p	the space of all sequences $\{x_n\}$ of real or complex numbers such that $\sum_{n=1}^{\infty} \lvert x_n\rvert^p < \infty$ for all $p \geq 1$
s	the space of all sequences of real or complex numbers
c	the space of all convergent sequences of real or complex numbers
$B[a,b]$	the space of all bounded real-valued functions defined on $[a,b]$
$C[a,b]$	the space of all continuous real-valued functions defined on $[a,b]$
$\rho(A,B)$	the distance between sets A and B
$\delta(A)$	the diameter of the set A
$S_r(x)$	the open sphere (or open ball) with center on x and radius r
$S_r[x]$	the closed sphere (or closed ball) with center on x and radius r
A°	the interior of the set A
int$A$	the interior of the set A
$\overline{A}$	the closure of the set A
clA	the closure of the set A
cl_XA	the closure of the set A in X
A^c	the complement of the set A
$b(A)$	the boundary of the set A
$\emptyset$	the empty set
$\langle x,y\rangle$	the scalar product of x and y
$\not\to$	does not converge
2^X	the family of all subsets of X
2^X_{cl}	the family of all nonempty closed and bounded subsets of X
2^X_q	the family of all nonempty compact subsets of X

$graph(f)$	the graph of a single-valued map f
$dom(f)$	the domain of a single-valued map f
$Graph(F)$	the graph of a set-valued map F
$Im(F)$	the image of of a set-valued map F
$Dom(F)$	the domain of a set-valued map F

Acronyms

EP	Equilibrium Problem
EPs	Equilibrium Problems
EVP	Ekeland's Variational Principle
COP	Constrained Optimization Problem
OP	Optimization Problem
MP	Minimization Problem
SPP	Saddle Point Problem
NEP	Nash Equilibrium Problem
FPP	Fixed Point problem
VIP	Variational Inequality Problem

Contents

Preface ..... *vii*
Notations and Abbreviations ..... *ix*

1 Basic Concepts **1**
1.1 Definition and Examples of Metric Spaces **1**
1.2 Distance between Sets and Diameter of a Set 13
1.3 Open Sets and Interior Points 15
1.4 Closed Sets and Closure of Sets 23
1.5 Subspaces 29

2 Complete Metric Spaces **33**
2.1 Convergent Sequences 33
2.2 Cauchy Sequences 37
2.3 Complete Metric Spaces 42
2.4 Completion 53

3 Separable Spaces **55**
3.1 Countability 55
3.2 Dense Sets 57
3.3 Nowhere Dense Sets 63

4 Compact Spaces **67**
4.1 Definitions and Basic Concepts 67
4.2 Sequentially Compact Spaces 74
4.3 Totally Bounded Spaces 77

5 Continuous Functions **87**
5.1 Definition and Characterizations 87
5.2 Continuous Functions and Compact Spaces 93
5.3 Uniform Continuous Functions 94
5.4 Homeomorphism and Equivalent Metrics 97
5.5 Uniform Convergence of Sequences of Functions 103

6 Connected Spaces **109**
6.1 Separated Sets 109
6.2 Connected Sets 111
6.3 Continuous Functions and Connected Sets 116
6.4 Components 118

7 Fixed Point Theorems **121**
7.1 Banach Contraction Theorem and its Applications 121
7.1.1 Application to System of Linear Equations 127
7.1.2 Application to Differential Equations 129
7.1.3 Application to Integral Equations 131
7.2 Further Extension of Banach Contraction Theorem 132
7.3 Caristi's Fixed Point Theorem 136

8 Set-Valued Maps **139**
8.1 Basic Concepts and Definitions 140
8.2 Continuity of Set-Valued Maps 145
8.3 Nadler's Fixed Point Theorem 152
8.4 Some Fixed Point Theorems for Set-Valued Maps 159

9 Ekeland's Variational Principle and its Applications **161**
9.1 Ekeland Variational's Principle in Complete Metric Spaces 161
9.2 Applications to Fixed Point Theorems 168
9.3 Applications to Optimization 171
9.4 Equilibrium Problems and Extended Ekeland's Variational Principle . 174
9.4.1 Equilibrium Problems 174
9.4.2 Extended Ekeland's Variational Principle 175

A Some Basic Inequalities **183**

B Partial Ordering **187**

C Nested Interval Property **189**
References 192

Index 195

Chapter 1
Basic Concepts

In our daily life, we observe that the distance between any two points should always be positive, it vanishes if and only if two points lie on the same position, the distance between two points is the same from whichever point it is measured, and the distance from one point to another cannot exceed sum of the distances from these points to an intermediate point. By exploiting these fundamental properties of a distance, we define the distance between two elements of a nonempty set.

1.1 Definition and Examples of Metric Spaces

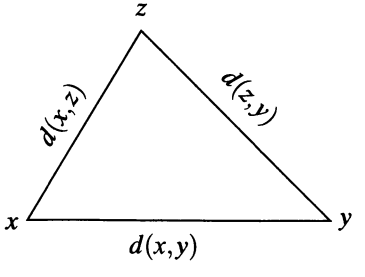
Definition 1.1. Let X be a nonempty set. A real-valued function d defined on $X \times X$ is said to be a *metric* on X if it satisfies the following conditions:

(M1) $d(x,y) \geq 0$ for all $x,y \in X$;
(M2) $d(x,y) = 0$ if and only if $x = y$;
(M3) $d(x,y) = d(y,x)$ for all $x,y \in X$; (symmetry)
(M4) $d(x,y) \leq d(x,z) + d(z,y)$ for all $x,y,z \in X$. (triangle inequality)

The set X together with a given metric d on X is called a *metric space* and it is denoted by (X,d). If there is no confusion likely to occur we, sometimes, denote the metric space (X,d) by X.

Since d is the generalization of the distance function, we interpret $d(x,y)$ as the distance between two elements x and y of the set X. Sometimes d is called a distance function on X.

The triangle inequality may be interpreted as that "the length of one side of a triangle can not exceed the sum of the length of the other two sides". Equivalently, the distance from x to y via any intermediate point z can not be shorter than the direct distance from x to y.



The metric d has the following properties:

1. From the triangle inequality (M4), it follows by induction that for any $x, y, z_1, z_2, \ldots, z_n \in X$,

$$\begin{aligned} d(x,y) &\le d(x,z_1) + d(z_1,y) \\ &\le d(x,z_1) + d(z_1,z_2) + d(z_2,y) \\ &\le d(x,z_1) + d(z_1,z_2) + d(z_2,z_3) + d(z_3,y) \\ &\quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\le d(x,z_1) + d(z_1,z_2) + \cdots + d(z_n,y). \end{aligned}$$

2. *For any $x, y, z \in X$, we have*

$$|d(x,z) - d(y,z)| \le d(x,y).$$

Indeed, it follows from (M3) and (M4) that

$$d(x,z) \le d(x,y) + d(y,z)$$

and

$$d(y,z) \le d(y,x) + d(x,z) = d(x,y) + d(x,z).$$

Thus

$$-d(x,y) \le d(x,z) - d(y,z) \le d(x,y),$$

and therefore,

$$|d(x,z) - d(y,z)| \le d(x,y).$$

Examples of Metric Spaces

Example 1.1. Let $X = \mathbb{R}$, the set of all real numbers. For any $x, y \in X$, define

$$d(x,y) = |x - y|.$$

Then (X,d) is a metric space and the metric d is called the *usual metric on* $\mathbb{R}$.

Verification. For all $x, y, z \in X$, we have

$$\begin{aligned} &\text{(M1)} && d(x,y) \ge 0; \\ &\text{(M2)} && d(x,y) = |x-y| = 0 \text{ if and only if } x = y; \\ &\text{(M3)} && d(x,y) = |x-y| = |-(x-y)| = |y-x| = d(y,x); \\ &\text{(M4)} && d(x,y) = |x-y| = |(x-z) + (z-y)| \\ & && \quad \le |x-z| + |z-y| \\ & && \quad = d(x,z) + d(z,y). \end{aligned}$$

In the verification of (M4) in Example 1.1, we used the fact that $|a+b| \le |a| + |b|$ for real numbers a and b. This property is also true for complex numbers a and b. Hence, we have the following example.

Example 1.2. Let $X = \mathbb{C}$, the set of all complex numbers. For any $x,y \in X$, define

$$d(x,y) = |x-y|.$$

Then (X,d) is a metric space and the metric d is called the *usual metric on* $\mathbb{C}$.

Example 1.3. Let X be any nonempty set. For any $x,y \in X$, define

$$d(x,y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y. \end{cases}$$

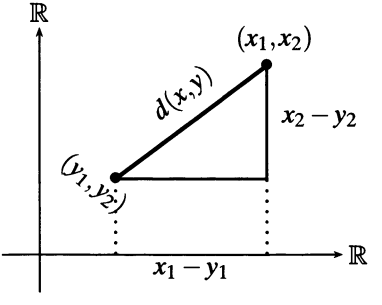
Then (X,d) is a metric space. The metric d is called *discrete metric* and the space (X,d) is called *discrete metric space*.

Remark 1.1. Example 1.3 shows that on each nonempty set, we can always define at least one metric that is discrete metric.

Example 1.4. Let $X = \mathbb{R}^2$, the set of all points in the coordinate plane. For any $x = (x_1,x_2)$, $y = (y_1,y_2)$ in X, define

$$d(x,y) = \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2}.$$

Then (X,d) is a metric space and $d(x,y)$ is the natural distance between two points in a plane. d is known as the *usual metric* on $\mathbb{R}^2$.



Verification. For any $x = (x_1,x_2)$, $y = (y_1,y_2)$ and $z = (z_1,z_2)$ in X. Obviously, $d(x,y) \geq 0$ and so (M1) holds.
(M2)

$$\begin{aligned} d(x,y) = 0 &\Leftrightarrow \sqrt{(x_1-y_1)^2 + (x_2-y_2)^2} = 0 \\ &\Leftrightarrow (x_1-y_1)^2 = 0 \text{ and } (x_2-y_2)^2 = 0 \\ &\Leftrightarrow x_1 = y_1 \text{ and } x_2 = y_2 \\ &\Leftrightarrow x = y. \end{aligned}$$

(M3)

$$\begin{aligned} d(x,y) &= \sqrt{(x_1-y_1)^2+(x_2-y_2)^2} \\ &= \sqrt{(y_1-x_1)^2+(y_2-x_2)^2} \\ &= d(y,x). \end{aligned}$$

(M4)

$$\begin{aligned} [d(x,y)]^2 &= (x_1-y_1)^2+(x_2-y_2)^2 \\ &= [(x_1-z_1)+(z_1-y_1)]^2+[(x_2-z_2)+(z_2-y_2)]^2 \\ &= (x_1-z_1)^2+(z_1-y_1)^2 \\ &\quad +2\left[\underbrace{(x_1-z_1)}_{a}\underbrace{(z_1-y_1)}_{b}+\underbrace{(x_2-z_2)}_{c}\underbrace{(z_2-y_2)}_{d}\right] \\ &\quad +(x_2-z_2)^2+(z_2-y_2)^2 \end{aligned}$$

Taking $x_1-z_1=a$, $z_1-y_1=b$, $x_2-z_2=c$ and $z_2-y_2=d$, and since

$$(ab+cd)^2 \le (a^2+c^2)(b^2+d^2),$$

we have

$$\begin{aligned} [d(x,y)]^2 &\le (x_1-z_1)^2+(x_2-z_2)^2 \\ &\quad +2\sqrt{(x_1-z_1)^2+(x_2-z_2)^2}\sqrt{(z_1-y_1)^2+(z_2-y_2)^2} \\ &\quad +(z_1-y_1)^2+(z_2-y_2)^2 \\ &= [d(x,z)]^2+2d(x,z)\,d(z,y)+[d(z,y)]^2 \\ &= [d(x,z)+d(z,y)]^2. \end{aligned}$$

Therefore,

$$d(x,y) \le d(x,z)+d(z,y).$$

Example 1.5. Let $X=\mathbb{R}^2$. For any $x=(x_1,x_2)$, $y=(y_1,y_2)$ in X, define

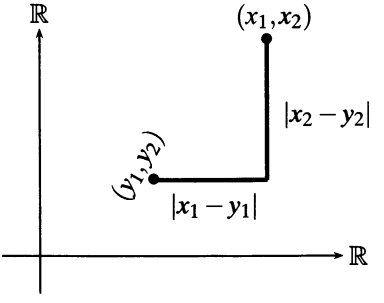
$$d(x,y) = |x_1-y_1|+|x_2-y_2|.$$

Then d is a metric on X and (X,d) is a metric space.

Verification. The conditions (M1), (M2) and (M3) are obvious. We prove only condition (M4). For any $x=(x_1,x_2)$, $y=(y_1,y_2)$ and $z=(z_1,z_2)$ in X,

$$\begin{aligned} d(x,y) &= |x_1-y_1|+|x_2-y_2| \\ &\le |x_1-z_1|+|z_1-y_1|+|x_2-z_2|+|z_2-y_2| \\ &= d(x,z)+d(z,y). \end{aligned}$$

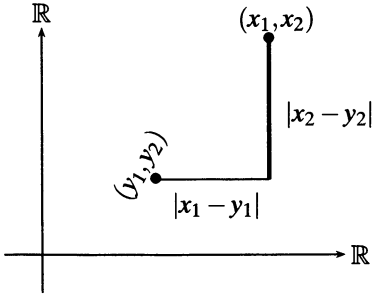
The distance between $x = (x_1, x_2)$ and $y = (y_1, y_2)$ with respect to the metric $d(x,y) = |x_1 - y_1| + |x_2 - y_2|$ is represented by the dark line in the following figure.



Example 1.6. Let $X = \mathbb{R}^2$. For any $x = (x_1, x_2)$, $y = (y_1, y_2)$ in X, define

$$d(x,y) = \max\{|x_1 - y_1|, |x_2 - y_2|\}.$$

Then d is a metric on X and (X,d) is a metric space. The distance between $x, y \in X$ is represented by the dark line in the following figure.



Remark 1.2. Examples 1.4, 1.5 and 1.6 show that more than one metric can be defined on a nonempty set.

Example 1.7. Let $X = \mathbb{R}^n$, the set of ordered n-tuples of real numbers. For any $x = (x_1, x_2, \ldots, x_n) \in X$ and $y = (y_1, y_2, \ldots, y_n) \in X$, we define

(a) $d_1(x,y) = \sum_{i=1}^{n} |x_i - y_i|$ (called taxicab metric)

(b) $d_2(x,y) = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{\frac{1}{2}}$, (called usual metric)

(c) $d_p(x,y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{\frac{1}{p}}$, $p \geq 1$

(d) $d_\infty(x,y) = \max_{1 \leq i \leq n} \{|x_i - y_i|\}$. (called max metric)

Verification. In view of Examples 1.4, 1.5 and 1.6, it is easy to verify that d_1, d_p and d_∞ are metrics on X.

The triangular inequality (M4) in the case of d_p requires the use of Minkowski inequality (Theorem A.3)

$$\begin{aligned} d_p(x,y) &= \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{\frac{1}{p}} = \left(\sum_{i=1}^{n} |x_i - z_i + z_i + y_i|^p\right)^{\frac{1}{p}} \\ &\leq \left(\sum_{i=1}^{n} |x_i - z_i| + |z_i - y_i|^p\right)^{\frac{1}{p}} \\ &\leq \left(\sum_{i=1}^{n} |x_i - z_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} |z_i - y_i|^p\right)^{\frac{1}{p}} \\ &= d_p(x,z) + d_p(z,y). \end{aligned}$$

Remark 1.3. Let $X = \mathbb{C}^n$, the set of ordered n-tuples of complex numbers. We can define the metrics d_1, d_p and d_∞ on X in the same way as in Example 1.7.

Example 1.8. Let ℓ^∞ be the space of all bounded sequences of real or complex numbers, that is,

$$\ell^\infty = \left\{ \{x_n\} \subseteq \mathbb{R} \text{ or } \mathbb{C} : \sup_{1 \leq n < \infty} |x_n| < \infty \right\}.$$

For any $x = \{x_n\} \in \ell^\infty$ and $y = \{y_n\} \in \ell^\infty$, define

$$d_\infty(x,y) = \sup_{1 \leq n < \infty} |x_n - y_n|.$$

Then it is easy to verify that d_∞ is a metric on ℓ^∞ and (ℓ^∞, d_∞) is a metric space.

Example 1.9. Let s be the space of all sequences of real or complex numbers, that is,

$$s = \{\{x_n\} \subseteq \mathbb{R} \text{ or } \mathbb{C}\}.$$

For any $x = \{x_n\}$ and $y = \{y_n\}$ in s, define

$$d(x,y) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{|x_n - y_n|}{1 + |x_n - y_n|}.$$

Then (s,d) is a metric space.

Verification. The series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{2^n} \frac{|x_n - y_n|}{1 + |x_n - y_n|}$ is convergent since its nth term is less than $\frac{1}{2^n}$. The conditions (M1), (M2) and (M3) can be easily verified.

Let $x = \{x_n\}$, $y = \{y_n\}$ and $z = \{z_n\}$ be in s. Then by the triangular inequality, we have

$$|x_n - y_n| \leq |x_n - z_n| + |z_n - y_n|$$

and hence[1]

$$\begin{aligned}\frac{|x_n-y_n|}{1+|x_n-y_n|} &\leq \frac{|x_n-z_n|+|z_n-y_n|}{1+|x_n-z_n|+|z_n-y_n|}\\ &= \frac{|x_n-z_n|}{1+|x_n-z_n|+|z_n-y_n|}+\frac{|z_n-y_n|}{1+|x_n-z_n|+|z_n-y_n|}\\ &\leq \frac{|x_n-z_n|}{1+|x_n-z_n|}+\frac{|z_n-y_n|}{1+|z_n-y_n|}.\end{aligned}$$

Multiplying both sides by $\frac{1}{2^n}$ and summing with respect to n, we obtain

$$d(x,y) \leq d(x,z)+d(z,y).$$

Example 1.10. Let $1 \leq p < \infty$. Consider the space ℓ^p of all sequences $\{x_n\}$ of real or complex numbers such that $\sum_{n=1}^{\infty} |x_n|^p < \infty$. Let $x = \{x_n\}$ and $y = \{y_n\} \in \ell^p$, we define

$$d(x,y) = \left(\sum_{n=1}^{\infty} |x_n-y_n|^p\right)^{\frac{1}{p}}.$$

Then d is a metric on ℓ^p and (ℓ^p, d) is a metric space.

Verification. The conditions (M1), (M2) and (M3) can be easily verified. Let $x = \{x_n\}, y = \{y_n\}$ and $z = \{z_n\}$ be sequences in ℓ^p. Then

$$\begin{aligned}d(x,y) &= \left(\sum_{n=1}^{\infty} |x_n-y_n|^p\right)^{\frac{1}{p}}\\ &= \left(\sum_{n=1}^{\infty} |x_n-z_n+z_n-y_n|^p\right)^{\frac{1}{p}}\\ &\leq \left(\sum_{n=1}^{\infty} |x_n-z_n|^p\right)^{\frac{1}{p}}+\left(\sum_{n=1}^{\infty} |z_n-y_n|^p\right)^{\frac{1}{p}}\\ &\quad \text{(by Minkowski's inequality)}\\ &= d(x,z)+d(z,y).\end{aligned}$$

Example 1.11. Let $B[a,b]$ be the space of all bounded real-valued functions defined on $[a,b]$, that is, $B[a,b] = \{f : [a,b] \to \mathbb{R} : f(t) \leq k \text{ for all } t \in [a,b]\}$. For $f, g \in B[a,b]$, we define

[1] Let $0 \leq \alpha \leq \beta$. Then $\alpha+\alpha\beta \leq \beta+\alpha\beta$. Dividing both sides by $(1+\alpha)(1+\beta)$, we have

$$\frac{\alpha}{1+\alpha} \leq \frac{\beta}{1+\beta}.$$

$$d(f,g) = \sup_{t\in[a,b]} |f(t) - g(t)|.$$

Then $(B[a,b],d)$ is a metric space.

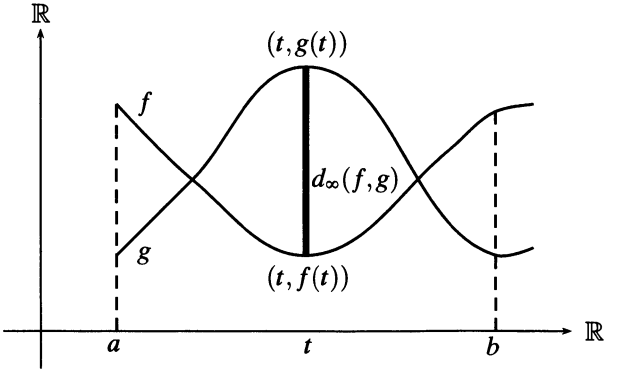
Example 1.12. Let $C[a,b]$ be the space of all continuous real-valued functions defined on $[a,b]$. For any $f,g \in C[a,b]$, we define the real-valued functions d_∞ and d_1 on $C[a,b] \times C[a,b]$ as follows:

$$d_\infty(f,g) = \max_{t\in[a,b]} |f(t) - g(t)|$$

and

$$d_1(x,y) = \int_a^b |f(t) - g(t)|\,dt,$$

where the integral is the Riemann integral which is possible because the functions f and g are continuous on $[a,b]$. Then d_∞ and d_1 are metrics on $C[a,b]$.

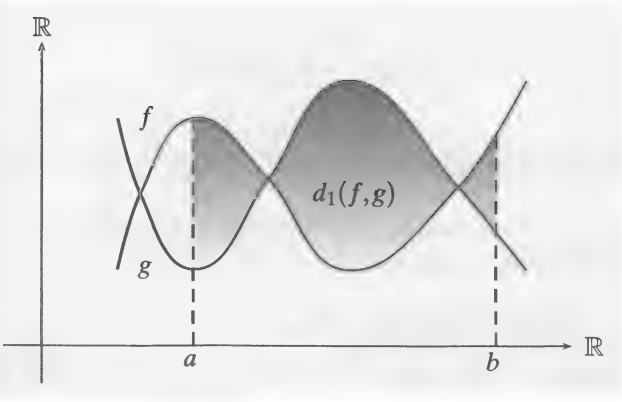


The metric d_∞ measures the distance from f to g as the maximum of vertical distances from points $(t,f(t))$ to $(t,g(t))$ on the graphs of f and g, respectively. $d_1(f,g)$ represents as a measure of the distance between the functions f and g to be the area enclosed between their graphs from $x = a$ to $x = b$.

We can also define another metric on $C[a,b]$. Let $f,g \in C[a,b]$, define

$$d(f,g) = \left(\int_a^b |f(t) - g(t)|^p\,dt\right)^{\frac{1}{p}}, \quad \text{for } p \geq 1.$$

Then $(C[a,b],d)$ is a metric space.



Remark 1.4. Let X be the set of all Riemann integrable functions on $[a,b]$. For $f,g \in X$, we define

$$d(f,g) = \int_a^b |f(x) - g(x)|dx.$$

Then d is not a metric on X. Indeed, if

$$f(x) = \begin{cases} 0 & \text{if } x = a \\ 1 & \text{if } a < x \leq b \end{cases}$$

and $g(x) = 1$ for all $x \in [a,b]$, then $d(f,g) = 0$. However, $f \neq g$. It may be noted that all other properties of a metric do hold. Such real-valued functions are called *pseudometrics* on X.

Definition 1.2. Let X be a nonempty set. A real-valued function $d : X \times X \to \mathbb{R}_+$ is said to be a *pseudometric* on X if it satisfies the following conditions:

(PM1) $d(x,y) \geq 0$ for all $x,y \in X$;
(PM2) $d(x,y) = 0$ if $x = y$;
(PM3) $d(x,y) = d(y,x)$ for all $x,y \in X$; (symmetry)
(PM4) $d(x,y) \leq d(x,z) + d(z,y)$ for all $x,y,z \in X$. (triangle inequality)

The set X together with a given pseudometric d on X is called a *pseudometric space*.

Example 1.13. Let $X = \mathbb{R}^2$, $x = (x_1,x_2)$, $y = (y_1,y_2) \in X$ and $d(x,y) = |x_1 - y_1|$. Then d is not a metric on X, however, it is a pseudometric on X. Indeed, for $x = (0,0)$, $y = (0,1) \in X$, we have $d(x,y) = 0$ but $x \neq y$. Therefore, it is not a metric on X. It can be easily checked that d satisfies the conditions **(PM1) – (PM4)**.

Exercise 1.1. Let (X,d) be a metric space. If $x,y,z,w \in X$, then prove that

$$|d(x,y) - d(z,w)| \leq d(x,z) + d(y,w).$$

Proof. Since

$$|d(x,z) - d(z,y)| \leq d(x,y),$$

we have

$$|d(x,y) - d(y,z)| \leq d(x,z)$$

and

$$|d(y,z) - d(z,w)| \leq d(y,w).$$

Hence

$$|d(x,y) - d(z,w)| \leq |d(x,y) - d(y,z)| + |d(y,z) - d(z,w)| \leq d(x,z) + d(y,w).$$

Exercise 1.2. Let X be a nonempty set and $d : X \times X \to \mathbb{R}$ be a real-valued function such that the following conditions hold: For all $x,y,z \in X$,

(i) $d(x,y) = 0$ if and only if $x = y$;
(ii) $d(x,y) = d(y,x)$ for all $x,y \in X$;
(iii) $d(x,y) \leq d(x,z) + d(z,y)$ for all $x,y,z \in X$.

Prove that d is a metric on X.
Hint: Put $y = x$ in (iii), we get $0 = d(x,x) \leq d(x,z) + d(z,x)$. By (ii), $0 \leq 2d(x,z)$ and so $d(x,z) \geq 0$ for all $x,z \in X$.

Exercise 1.3. Let X be a nonempty set and $d : X \times X \to \mathbb{R}$ be a real-valued function such that the following conditions hold: For all $x,y,z \in X$,

(i) $d(x,y) = 0$ if and only if $x = y$;
(ii) $d(x,y) \leq d(z,x) + d(z,y)$ for all $x,y,z \in X$.

Prove that d is a metric on X.

Exercise 1.4. Let X be a nonempty set and $d : X \times X \to \mathbb{R}$ be a real-valued function such that the following conditions hold: For all $x,y,z \in X$,

(i) $d(x,y) = 0$ if and only if $x = y$;
(ii) $d(x,y) \leq d(x,z) + d(z,y)$ for all $x,y,z \in X$.

Show that these conditions are not sufficient to make the function d a metric on the set X.

Exercise 1.5. Let $X = \mathbb{R}^2$ and $x = (0,2)$, $y = (3,6)$ in X. Find the distance between x and y by using the metrics of Examples 1.4, 1.5 and 1.6.

Exercise 1.6. Let c be the space of all convergent sequences of real or complex numbers. For $x = \{x_n\}$ and $y = \{y_n\}$ in c, let

$$d(x,y) = \sup_{1 \leq n < \infty} |x_n - y_n|.$$

Prove that d is a metric on c.

Exercise 1.7. Let $\mathbb{K}$ be the set of all real or complex numbers. Prove that for each $x,y \in \mathbb{K}$,

$$d_1(x,y) = \min\{1, |x-y|\}$$

and

$$d(x,y) = \begin{cases} 0 & \text{if } x = y \\ |x| + |y| & \text{if } x \neq y \end{cases}$$

are metrics on $\mathbb{K}$.

Exercise 1.8. Let $X = [0,1)$. For each $x,y \in X$, let

$$d(x,y) = |x-y|.$$

Prove that d is a metric on X.

Exercise 1.9. Let $X = \mathbb{Q}$, the set of all rational numbers. Show that for each $x,y \in X$,

$$d(x,y) = |x-y|$$

is a metric on X.

Exercise 1.10. Let $X = \{x \in \mathbb{R} : x > 0\}$ and $d(x,y) = \left|\frac{1}{x} - \frac{1}{y}\right|$ for all $x,y \in X$. Show that d is a metric on X.

Exercise 1.11. Let $X = \mathbb{R}$ and $d : X \times X \to \mathbb{R}$ be a real-valued function defined by $d(x,y) = \sqrt{|x-y|}$ for all $x,y \in X$. Show that d is metric on X.
Hint: For any $x,y,z \in X$, we have

$$\begin{aligned}[d(x,y)]^2 &= |x-y| \le |x-z| + |z-y| \\ &= [d(x,z)]^2 + [d(z,y)]^2 \\ &\le [d(x,z) + d(z,y)]^2.\end{aligned}$$

Exercise 1.12. Let $X = \mathbb{Q}$, the set of all rational numbers, for all $\frac{p}{q}, \frac{r}{s} \in \mathbb{Q}$ written in their lowest terms, $q,s > 0$, and $d\left(\frac{p}{q}, \frac{r}{s}\right) = \left|\frac{1}{q} - \frac{1}{s}\right|$. Is d a metric on X?
Hint: Consider $\frac{1}{3}, \frac{2}{3} \in \mathbb{Q}$ in their lowest terms. Then $d\left(\frac{1}{3}, \frac{2}{3}\right) = \left|\frac{1}{3} - \frac{1}{3}\right| = 0$ but $\frac{1}{3} \neq \frac{2}{3}$.

Exercise 1.13. Show that $X = \mathbb{R} \setminus \mathbb{Q}$ with $d(x,y) = \left|\frac{1}{x} - \frac{1}{y}\right|$ is a metric space.

Exercise 1.14. Show that $\mathbb{C}$ with

$$d(z_1,z_2) = \begin{cases} \min\{|z_1| + |z_2|, |z_1 - 1| + |z_2 - 1|\} & \text{if } z_1 \neq z_2 \\ 0 & \text{otherwise} \end{cases}$$

is a metric space.

Exercise 1.15. Determine whether the following real-valued functions defined on $\mathbb{R} \times \mathbb{R}$ are metrics on $\mathbb{R}$. For all $(x,y) \in \mathbb{R} \times \mathbb{R}$,

(a) $d_1(x,y) = [|x-y|]$, the greatest integer less than or equal to $|x-y|$;
(b) $d_2(x,y) = \ln|x-y|$;
(c) $d_3(x,y) = |\sin(x-y)|$;
(d) $d_4(x,y) = e^{|x-y|}$;
(e) $d_5(x,y) = |\cos(x-y)|$;
(f) $d_6(x,y) = |x^2 - y^2|$;
(g) $d_7(x,y) = 2|x-y|$;
(h) $d_8(x,y) = (x-y)^3$;
(i) $d_9(x,y) = |x-y|^3$.

Exercise 1.16. Let $X = \mathbb{R}$ and $d : X \times X \to \mathbb{R}$ be a real-valued function defined by

$$d(x,y) = \begin{cases} |x| + |x-y| + |y| & \text{if } x \neq y, \\ 0 & \text{otherwise.} \end{cases}$$

Show that (X,d) is a metric space.

Exercise 1.17. Let $X = \mathbb{R}^2$ and for each $x = (x_1,x_2)$, $y = (y_1,y_2) \in X$, let

$$d(x,y) = \begin{cases} |x_1 - y_1| & \text{if } x_2 = y_2 \\ |x_1| + |y_1| + |x_2 - y_2| & \text{if } x_2 \neq y_2. \end{cases}$$

Then prove that (X,d) is a metric space.

Exercise 1.18. Let $X = \mathbb{R}^2$ and $d : X \times X \to \mathbb{R}$ be a real-valued function defined by

$$d(x,y) = d((x_1,x_2),(y_1,y_2)) = \begin{cases} |x_1 - y_1| & \text{if } x_2 = y_2 \\ |x_1| + |x_2 - y_2| + |y_1| & \text{otherwise.} \end{cases}$$

Show that (X,d) is a metric space.
Hint: Let $x = (x_1,x_2)$, $y = (y_1,y_2)$ and $z = (z_1,z_2)$ be in X. Observe that $d(x,y) \geq |x_1 - y_1|$. If $x_2 = y_2$, then $d(x,y) = |x_1 - y_1| \leq |x_1 - z_1| + |z_1 - y_1| \leq d(x,z) + d(z,y)$. If $x_2 \neq y_2$, then z_2 cannot be equal to both x_2 and y_2; so assume that $z_2 \neq x_2$. Then

$$\begin{aligned} d(x,y) &= |x_1| + |x_2 - y_2| + |y_1| \leq |x_1| + |x_2 - z_2| + |z_2 - y_2| + |y_1| \\ &\leq \begin{cases} (|x_1| + |x_2 - z_2| + |z_1|) + |z_1 - y_1| & \text{if } y_2 = z_2, \\ (|x_1| + |x_2 - z_2| + |z_1|) + (|z_1| + |z_2 - y_2| + |y_1|) & \text{if } y_2 \neq z_2 \end{cases} \\ &= d(x,z) + d(z,y). \end{aligned}$$

Exercise 1.19. Let $X = \mathbb{R}^2$ and for all $x = (x_1,x_2)$, $y = (y_1,y_2) \in X$, $d(x,y) = \min\{|x_1 - y_1|, |x_2 - y_2|\}$. Is d a metric on X?
Hint: Choose $x = (2,3)$, $y = (2,4)$. Then $x \neq y$, but $d(x,y) = \min\{0,1\} = 0$. Hence d is not a metric on X.

Exercise 1.20. Show that $X = \mathbb{R}^2$ with $d(x,y) = d((x_1,y_1),(x_2,y_2)) = |x_2 - y_2|$ is not a metric space.

Exercise 1.21. Let $X = C[0,1]$ and $d_\infty : X \times X \to \mathbb{R}$ be a real-valued function defined by $d_\infty(f,g) = \sup_{t \in [0,1]} |f(t) - g(t)|$ for all $f, g \in C[0,1]$. Then calculate the distance between $f(t) = t$ and $g(t) = t^2$.
Hint: $f(t) - g(t) = t - t^2 = -(\frac{1}{2} - t)^2 + \frac{1}{4}$ and so $\sup_{t \in [0,1]} |f(t) - g(t)| = \frac{1}{4}$.

Exercise 1.22. Let d_1 and d_2 be metrics on a set X. Is $\min\{d_1,d_2\}$ also a metric on X? Justify your answer.

Exercise 1.23. Let (X_i,d_i), $i = 1,2,\ldots,n$, be metric spaces and $X = X_1 \times X_2 \times \cdots \times X_n$. Then prove that for each $x = (x_1,x_2,\ldots,x_n) \in X$ and $y = (y_1,y_2,\ldots,y_n) \in X$,

$$d_1(x,y) = \max_{1\le i\le n} d_i(x_i,y_i)$$

and

$$d_2(x,y) = \sum_{i=1}^{n} d_i(x_i,y_i)$$

are metrics on X.

Exercise 1.24. Let (X,d) be a metric space. Prove that for each $x,y \in X$,

$$d^*(x,y) = \frac{d(x,y)}{1+d(x,y)}$$

is also a metric on X.
Hint: Use $\dfrac{a}{1+a} + \dfrac{b}{1+b} \ge \dfrac{a+b}{1+a+b}$ for all $a \ge 0, b \ge 0$.

1.2 Distance between Sets and Diameter of a Set

Definition 1.3. Let (X,d) be a metric space and let A and B be nonempty subsets of X. The *distance between the sets A and B*, denoted by $\rho(A,B)$, is given by

$$\rho(A,B) = \inf\{d(x,y) : x \in A, y \in B\}.$$

Since $d(x,y) = d(y,x)$, we have $\rho(A,B) = \rho(B,A)$.

If A consists of a single point x, then

$$\rho(\{x\},B) = \inf\{d(x,y) : y \in B\}.$$

It is called the *distance of a point $x \in X$ from the set B*, and is denoted by $\rho(x,B)$.

Remark 1.5. (i) The equation $\rho(x,B) = 0$ does not imply that x belongs to B.
(ii) If $\rho(A,B) = 0$, then it does not imply that A and B have common points.

Example 1.14. Let $A = \{x \in \mathbb{R} : x > 0\}$ and $B = \{x \in \mathbb{R} : x < 0\}$ be subsets of $\mathbb{R}$ with the usual metric. Then $\rho(A,B) = 0$, but A and B have no common point. If $x = 0$ then $\rho(x,B) = 0$, but $x \notin B$.

Definition 1.4. Let (X,d) be a metric space and A be a nonempty subset of X. The *diameter of A*, denoted by $\delta(A)$, is given by

$$\delta(A) = \sup\{d(x,y) : x,y \in A\}.$$

The set A is called *bounded* if $\delta(A) \le k < \infty$. In other words, A is bounded if its diameter is finite, otherwise it is called *unbounded.*

In particular, the metric space (X,d) is bounded if the set X is bounded.

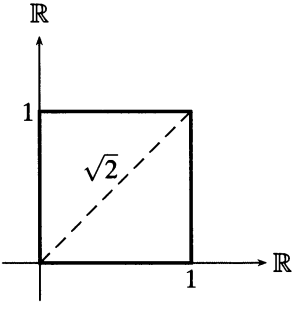
Example 1.15. (a) The real line with the usual metric is an unbounded metric space.

(b) In $\mathbb{R}$ with the usual metric, the intervals $[a,b]$, (a,b), $[a,b)$ and $(a,b]$ are bounded. But $[a,\infty)$ and $(-\infty,a]$ are not bounded.

(c) The space s of all sequences of real or complex numbers with the metric defined in Example 1.9 is a bounded space since $d(x,y) < \sum_{n=1}^{k} \frac{1}{2^n}$.

(d) Every set in a discrete metric space (X,d) is bounded and its diameter is 1.

Example 1.16. Consider the set $S = \{(x,y) \in \mathbb{R} \times \mathbb{R} : 0 \le x \le 1,\ 0 \le y \le 1\}$ in $\mathbb{R}^2$. With the usual metric d, the diameter of S is $\sqrt{2}$; with the taxicab metric, its diameter is 2; with the max metric, its diameter is 1; and with the discrete metric its diameter is 1.



Remark 1.6. Let (X,d) be a metric space. We can define other metrics on X with the help of the metric d in the following manner:

$$d_1(x,y) = \frac{d(x,y)}{1+d(x,y)} \quad \text{and} \quad d_2(x,y) = \min\{1, d(x,y)\}.$$

Then d_1 and d_2 are metrics on X and with these metrics (X,d_1) and (X,d_2) are bounded metric spaces irrespective of whether the metric space (X,d) is bounded or not.

Exercise 1.25. Determine the distance from $(3,4)$ to the set $[0,1] \times [0,1]$ in $\mathbb{R}^2$ with respect to the metrics (a) usual, (b) taxicab, (c) max, and (d) discrete.

Exercise 1.26. Let $A = \{x = (x_1,x_2,x_3) \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 \le 1\}$ be a set in $\mathbb{R}^3$. Compute the diameter of A with respect to each of the following metrics: (a) usual, (b) taxicab, (c) max, and (d) discrete.

Exercise 1.27. Let $A = \{x = (x_1,x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 \le 1\}$ and $x = (1,1)$. Find the distance from x to A for the metrics (a) usual, (b) taxicab, (c) max, and (d) discrete.

Exercise 1.28. Let A and B be nonempty subsets of a metric space (X,d). Prove that

(a) $\delta(A) = 0$ if and only if A is a singleton set;
(b) For each $x \in A$, $y \in B$, $\rho(A,B) \le d(x,y)$;
(c) If $A \subseteq B$, then $\delta(A) \le \delta(B)$;
(d) For each $x \in A$, $y \in B$, $d(x,y) \le \delta(A \cup B)$;
(e) $\delta(A \cup B) \le \delta(A) + \rho(A,B) + \delta(B)$;
(f) If $A \cap B \neq \emptyset$, then $\delta(A \cup B) \le \delta(A) + \delta(B)$;
(g) $d(x,A) \le d(x,y) + d(y,A)$ for all $x,y \in X$.

Proof ((e) and (f)). Let a and b be arbitrary elements of A and B, respectively, and let $x,y \in A \cup B$. If both x and y are in A, then $d(x,y) \leq \delta(A)$. If both x and y are in B, then $d(x,y) \leq \delta(B)$.

If $x \in A$ and $y \in B$, then by the triangle inequality, we have

$$\begin{aligned} d(x,y) &\leq d(x,a) + d(a,y) \\ &\leq d(x,a) + d(a,b) + d(b,y) \\ &\leq \delta(A) + d(a,b) + \delta(B). \end{aligned}$$

Similarly, if $x \in B$ and $y \in A$, we have

$$d(x,y) \leq \delta(A) + d(a,b) + \delta(B).$$

Thus,

$$d(x,y) \leq \delta(A) + d(a,b) + \delta(B) \quad \text{for all } x,y \in A \cup B.$$

Therefore,

$$\delta(A \cup B) \leq \delta(A) + d(a,b) + \delta(B) \quad \text{for all } a \in A,\ b \in B.$$

Hence,

$$\delta(A \cup B) \leq \delta(A) + \rho(A,B) + \delta(B).$$

Now, if $A \cap B \neq \emptyset$, we have $\rho(A,B) = 0$ and hence $\delta(A \cup B) \leq \delta(A) + \delta(B)$.

1.3 Open Sets and Interior Points

Definition 1.5. Let (X,d) be a metric space. Given a point $x_0 \in X$ and a real number $r > 0$, the sets

$$S_r(x_0) = \{y \in X : d(x_0,y) < r\}$$

and

$$S_r[x_0] = \{y \in X : d(x_0,y) \leq r\}$$

are called *open sphere* (or *open ball*) and *closed sphere* (or *closed ball*), respectively, with center x and radius r.

Remark 1.7. (i) The open and closed spheres are always nonempty, since $x_0 \in S_r(x_0) \subseteq S_r[x_0]$.

(ii) Every open (respectively, closed) sphere in $\mathbb{R}$ with the usual metric is an open (respectively, closed) interval. But the converse is not true; For example, $(-\infty,\infty)$ is an open interval in $\mathbb{R}$ but not an open sphere.

Example 1.17. (a) In the metric space $\mathbb{R}$ with the usual metric, the spheres $S_r(x_0)$ and $S_r[x_0]$ are intervals

$$(x_0 - r, x_0 + r) \quad \text{and} \quad [x_0 - r, x_0 + r],$$

respectively.

(b) In the metric space $\mathbb{C}$ with the usual metric, the sphere $S_r(z_0)$ and $S_r[z_0]$ are circular discs

$$|z-z_0| < r \quad \text{and} \quad |z-z_0| \leq r,$$

respectively, where $z_0 \in \mathbb{C}$ and $r > 0$.

(c) Let X be a nonempty set with the discrete metric d. Then the open sphere $S_r(x_0)$ is

$$S_r(x_0) = \begin{cases} \{x_0\} & \text{if } 0 < r \leq 1, \\ X & \text{if } r > 1, \end{cases}$$

and the closed sphere $S_r[x_0]$ is

$$S_r[x_0] = \begin{cases} \{x_0\} & \text{if } 0 < r < 1, \\ X & \text{if } r \geq 1. \end{cases}$$

(d) Let $X = [0,1)$ be a metric space with the usual metric $d(x,y) = |x-y|$ for all $x,y \in X$. Then the open sphere $S_r(0)$ is

$$S_r(0) = \begin{cases} [0,r) & \text{if } r \leq 1, \\ [0,1) & \text{if } r > 1, \end{cases}$$

and the closed sphere $S_r[0]$ is

$$S_r[0] = \begin{cases} [0,r) & \text{if } r < 1, \\ [0,1) & \text{if } r \geq 1. \end{cases}$$

(e) In $\mathbb{R}^2$, the open sphere with center 0 and radius 1 with respect to the metrics d_1, d_2 and d_∞, respectively, (in Example 1.7), are

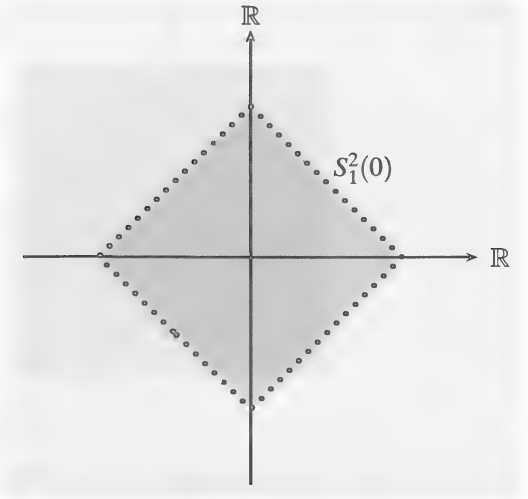
$$S_1^1(0) = \{y = (y_1,y_2) \in \mathbb{R}^2 : |y_1| + |y_2| < 1\},$$

$$S_1^2(0) = \left\{y = (y_1,y_2) \in \mathbb{R}^2 : \sqrt{y_1^2+y_2^2} < 1\right\},$$

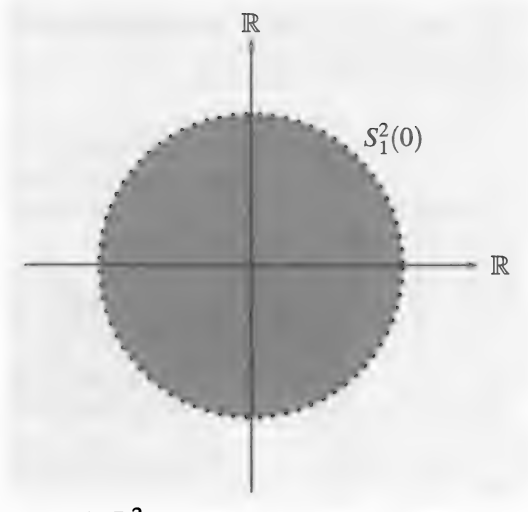
and

$$S_1^\infty(0) = \{y \in (y_1,y_2) \in \mathbb{R}^2 : \max(|y_1|,|y_2|) < 1\}.$$

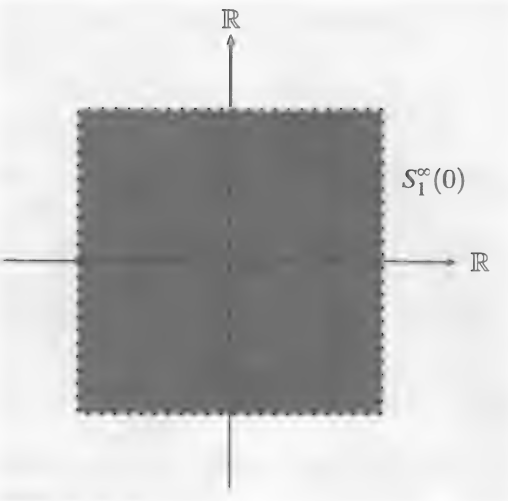
Similarly, we can define the closed spheres.



The sphere in $\mathbb{R}^2$ with respect to the metric
$d_1(x,y) = d_1((x_1,x_2),(y_1,y_2)) = |x_1 - y_1| + |x_2 - y_2|$



The sphere in $\mathbb{R}^2$ with respect to the metric
$d_2(x,y) = d_2((x_1,x_2),(y_1,y_2)) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$

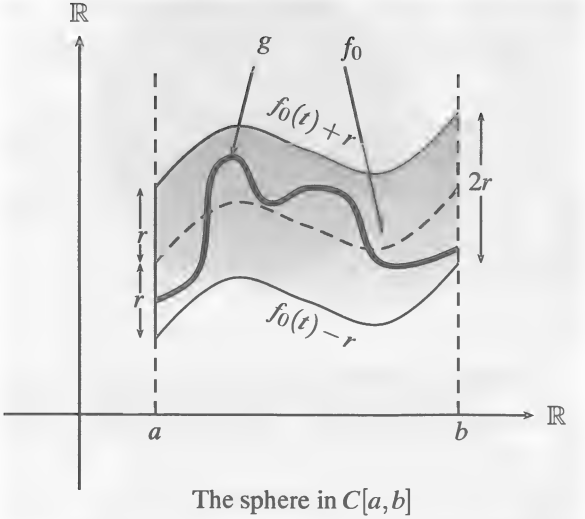


The sphere in $\mathbb{R}^2$ with respect to the metric
$d_\infty(x,y) = d_\infty((x_1,x_2),(y_1,y_2)) = \max\{|x_1-y_1|,|x_2-y_2|\}$

(f) In the metric space $C[a,b]$ with respect to the metric $d_\infty(f,g) = \max_{t\in[a,b]} |f(t)-g(t)|$, the open sphere $S_r(f_0)$ with center f_0 and radius r is the set of continuous functions g such that

$$\sup_{t\in[0,1]} |f_0(t)-g(t)| < r,$$

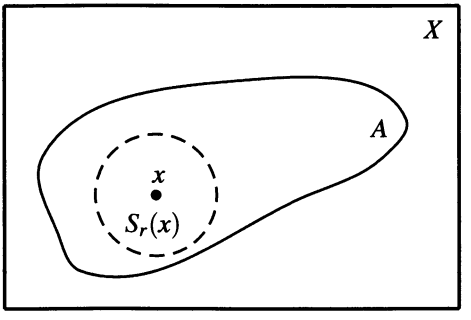
that is, the set of continuous functions g whose graphs lie within the shaded band of vertical width $2r$ centered on the graph of f_0.



The sphere in $C[a,b]$

Definition 1.6. Let A be a nonempty subset of a metric space X.

- A point $x \in A$ is said to be an *interior point* of A if x is the center of some sphere contained in A. In other words, $x \in A$ is an interior point of A if there exists $r > 0$ such that $S_r(x) \subseteq A$.



- The set of all interior points of A is called *interior* of A and it is denoted by A°, that is,
$$A^\circ = \{x \in A : S_r(x) \subseteq A \text{ for some } r > 0\}.$$
- The set A is said to be *open* if each of its points is the center of some open sphere contained entirely in A; That is to say, A is an open set if for each $x \in A$, there exists $r > 0$ such that $S_r(x) \subseteq A$.
- Let $x \in X$. The set A is said to be a *neighborhood* of x if there exists an open sphere centered at x and contained in A, that is, if $S_r(x) \subseteq A$ for some $r > 0$. In case, A is an open set, it is called an *open neighborhood* of x.

Remark 1.8. (i) In particular, an open sphere $S_r(x)$ with center x and radius r is a neighborhood of x.
(ii) The interior of A is the neighborhood of each of its points.
(iii) Every open set is the neighborhood of each of its points.
(iv) The set A is open if and only if each of its points is an interior point, that is, $A = A^\circ$.

Example 1.18. 1. Let $\mathbb{R}$ be the metric space with the usual metric and A be a subset of $\mathbb{R}$.

(a) $A = (a,b)$, $[a,b)$, $[a,b,]$, or $(a,b]$, then $A^\circ = (a,b)$
(b) If $A = \mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ or the set of irrational numbers, then $A^\circ = \emptyset$.
(c) If A is a finite set, then $A^\circ = \emptyset$.
(d) If $A = C$ the cantor set, then $A^\circ = \emptyset$.
(e) If $A = \bigcap_{n=1}^{\infty} [-3, \frac{1}{n})$, then $A^\circ = (-3,0)$.
(f) If $A = (0,1) \cap \mathbb{Q}$, then $A^\circ = \emptyset$.
(g) If $A = \emptyset$, then $A^\circ = \emptyset$.
(h) If $A = \mathbb{R}$, then $A^\circ = \mathbb{R}$.

2. Let A be a nonempty subset of a discrete metric space X. Then $A^\circ = A$.

Example 1.19. 1. In $\mathbb{R}$ with the usual metric

(a) $\mathbb{R}$ is an open set;
(b) (a,b) is an open set;
(c) $(a,b]$, $[a,b)$ and $[a,b]$ are not open sets;
(d) The set $\left\{1, \frac{1}{2}, \frac{1}{3}, \cdots\right\}$ is not open;
(e) A set consists a singleton is not an open set;
(f) The set of all rational numbers $\mathbb{Q}$ is not open. But it is open with respect to the metric $d(x,y) = |x-y|$ defined on $\mathbb{Q}$;
(g) The cantor set C is not an open set.

2. Let $X = [0,1)$ with the usual metric $d(x,y) = |x-y|$ for all $x,y \in X$. Then $[0,\alpha)$, $\alpha \leq 1$, is an open set.

3. In the discrete metric space X, every subset of X is an open set.

Remark 1.9. (i) In a metric space X, the empty set $\emptyset$ and the whole space X are open sets.
(ii) Whether a set is open or not depends upon the space in which it is considered. For example, identify the real line $\mathbb{R}$ with horizontal axis $\{(x,0) \in \mathbb{R}^2 : x \in \mathbb{R}\}$ in $\mathbb{R}^2$. $\mathbb{R}$ is not an open subset of $\mathbb{R}^2$ since $\mathbb{R}$ does not contain any open sphere in $\mathbb{R}^2$. However, $\mathbb{R}$ in the metric space $\mathbb{R}$ with the usual metric is an open set.

Theorem 1.1. *Let A and B be two subsets of a metric space X. Then*

(a) $A \subseteq B$ implies $A^\circ \subseteq B^\circ$;
(b) $(A \cap B)^\circ = A^\circ \cap B^\circ$;
(c) $(A \cup B)^\circ \supseteq A^\circ \cup B^\circ$.

Proof. (a) Let $x \in A^\circ$. Then there exists an open sphere $S_r(x) \subseteq A$ for some $r > 0$. Since $A \subseteq B$, $S_r(x) \subseteq B$ and hence $x \in B^\circ$. Thus $A^\circ \subseteq B^\circ$.

(b) Let $x \in (A \cap B)^\circ$. Then there exists an open sphere $S_r(x) \subseteq A \cap B$ for some $r > 0$. Therefore, $S_r(x) \subseteq A$ and $S_r(x) \subseteq B$ and hence $x \in A^\circ$ and $x \in B^\circ$. So, $x \in A^\circ \cap B^\circ$ and thus $(A \cap B)^\circ \subseteq A^\circ \cap B^\circ$.

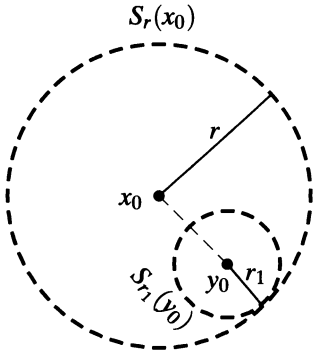
To prove the reverse inclusion, let us suppose that $y \in A^\circ \cap B^\circ$. Then $y \in A^\circ$ and $y \in B^\circ$ and therefore, there exist open spheres $S_{r_1}(y) \subseteq A$ and $S_{r_2}(y) \subseteq B$ for some $r_1, r_2 > 0$. Set $r = \min\{r_1, r_2\}$. Then $S_r(y) \subseteq A \cap B$ and hence $y \in (A \cap B)^\circ$. Consequently, $A^\circ \cap B^\circ \subseteq (A \cap B)^\circ$.

(c) Let $x \in A^\circ \cup B^\circ$. Then either $x \in A^\circ$ or $x \in B^\circ$. This implies that there exists an open sphere $S_r(x) \subseteq A$ or $S_r(x) \subseteq B$ for some $r > 0$. So, we have $S_r(x) \subseteq A \cup B$ and therefore $x \in (A \cup B)^\circ$. Hence $A^\circ \cup B^\circ \subseteq (A \cup B)^\circ$.

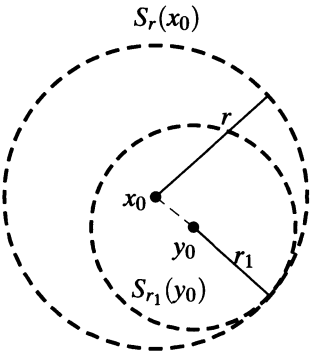
Remark 1.10. $(A \cup B)^\circ \not\subseteq A^\circ \cup B^\circ$. For example, let $X = \mathbb{R}$ be the usual metric space and $A = [0,1]$ and $B = [1,2]$. Then $A \cup B = [0,2]$. Note that $A^\circ = (0,1)$, $B^\circ = (1,2)$ and $(A \cup B)^\circ = (0,2)$. This shows that $A^\circ \cup B^\circ \subseteq (A \cup B)^\circ$ but $(A \cup B)^\circ \not\subseteq A^\circ \cup B^\circ$.

Theorem 1.2. *Let (X,d) be a metric space. Then*

(a) each open sphere in X is an open set;
(b) a subset A of X is open if and only if it is the union of open spheres.



$r_1 < d(x_0, y_0)$



$r_1 > d(x_0, y_0)$

Proof. (a) Let $S_r(x_0) = \{x \in X : d(x, x_0) < r\}$ be an open sphere in X and let $y_0 \in S_r(x_0)$. We have to produce an open sphere centered at y_0 and contained in $S_r(x_0)$. Since $y_0 \in S_r(x_0)$, we have $d(x_0, y_0) < r$. Set

$$r_1 = r - d(x_0, y_0) > 0.$$

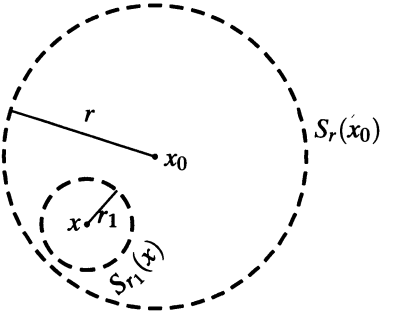
Consider

$$S_{r_1}(y_0) = \{y \in X : d(y, y_0) < r_1\}.$$

We have to show that $S_{r_1}(y_0) \subseteq S_r(x_0)$. For this, let $y \in S_{r_1}(y_0)$ be arbitrary. Then $d(y, y_0) < r_1$ and therefore

$$\begin{aligned} d(x_0, y) &\leq d(x_0, y_0) + d(y_0, y) \qquad \text{(by triangle inequality)} \\ &< d(x_0, y_0) + r_1 = r. \end{aligned}$$

Thus $y \in S_r(x_0)$ and consequently, $S_{r_1}(y_0) \subseteq S_r(x_0)$.



(b) Suppose that A is an open set. Then each of its points is the center of an open sphere contained in A. Hence A is the union of all the open spheres contained in it.

To prove the converse part, let us assume that A is the union of a collection $\mathscr{F}$ of open spheres. Let $x \in A$ be arbitrary. Then, x belongs to some open sphere, say

$S_r(x_0) \in \mathscr{F}$. Since each open sphere is an open set, x is the center of an open sphere $S_{r_1}(x)$ such that $S_{r_1}(x) \subseteq S_r(x_0)$. But $S_r(x_0) \subseteq A$ and hence $S_{r_1}(x) \subseteq A$. Therefore A is open

Theorem 1.3. *Let (X,d) be a metric space. Then*

(a) arbitrary union of open sets in X is open;
(b) finite intersection of open sets in X is open.

Proof. (a) Let Λ be any index set, $\{A_\alpha\}_{\alpha\in\Lambda}$ be a family of open sets in X and $A = \bigcup_{\alpha\in\Lambda} A_\alpha$. Since each A_α is open, it is the union of open spheres for each $\alpha \in \Lambda$. Then A is the union of unions of open spheres. Hence, by Theorem 1.2, A is open.

(b) Let $\{A_i : i = 1,2,\ldots,n\}$ be the finite family of open sets in X and let $A = \bigcap_{i=1}^{n} A_i$. Let $x \in A$. Then x is in each A_i. But each A_i is open, hence for each i, there exists $r_i > 0$ such that $S_{r_i}(x) \subseteq A_i$. Set $r = \min\{r_1, r_2, \ldots, r_n\}$. Then

$$S_r(x) \subseteq S_{r_i}(x) \subseteq A_i \quad \text{for each} \quad i = 1,2,\ldots,n.$$

Therefore, $S_r(x) \subseteq \bigcap_{i=1}^{n} A_i = A$ and hence A is open.

Remark 1.11. Arbitrary intersection of open sets need not be open. For example, let $X = \mathbb{R}$ with the usual metric. Consider the family $A_n = \left\{\left(-\frac{1}{n}, \frac{1}{n}\right) : n \in \mathbb{N}\right\}$ of open sets. Then $\bigcap_{i=1}^{\infty} A_n = \{0\}$ is not open.

Theorem 1.4. *Let A be a subset of a metric space X. Then A° is the largest open subset of A.*

Proof. First of all, we shall prove that A° is an open set. For that, let $x \in A^\circ$ be arbitrary. Then, by definition, there exists an open sphere $S_r(x) \subseteq A$ for some $r > 0$. But $S_r(x)$ is an open set, so each of its points is the center of some open sphere contained in $S_r(x)$. Therefore, each point of $S_r(x)$ is the interior point of A, that is, $S_r(x) \subseteq A^\circ$. Thus, x is the center of an open sphere contained in A°. Hence A° is an open set.

Let $B \subseteq A$ be an arbitrary open set and let $x \in B$. Then there exists $S_r(x) \subseteq B \subseteq A$ for some $r > 0$. This implies that $x \in A^\circ$ and hence $B \subseteq A^\circ \subseteq A$. Since A° is open, A° is the largest open subset of A.

Remark 1.12. A° is the union of all open subsets of A.

Exercise 1.29. Find the open spheres with center 0 and radius 1 in the metric spaces with respect to the metrics defined in Exercises 1.7 and 1.10.

Exercise 1.30. Let A be a subset of a metric space X. Prove that $(A^\circ)^\circ = A^\circ$.

Exercise 1.31. In $\mathbb{R}^n$, let R denote the set of points having only rational coordinates and I its complements, that is, the set of points having at least one irrational coordinate. Then prove that $R^\circ = I^\circ = \emptyset$.

Exercise 1.32. Let (X,d) be a metric space, $a \in X$ and $0 < r < r'$. Prove that the set $\{x \in X : r < d(x,a) < r'\}$ is open in X.

Exercise 1.33. Let (X,d) be a metric space and

$$d^*(x,y) = \frac{d(x,y)}{1+d(x,y)}.$$

Prove that the family of open sets with respect to the metric d is same as the family of open sets with respect to the metric d^*.

Exercise 1.34. Let R be the same as in Exercise 1.31. Prove that

(a) every nonempty open set in $\mathbb{R}^n$ contains a member of R;
(b) every nonempty open set in $\mathbb{R}^n$ contains infinitely many members of R.

Exercise 1.35. Let (X,d) be a metric space and x, y be distinct points of X. Prove that there exist disjoint open spheres centered on x and y.
Hint: Since $x \neq y$, $d(x,y) > 0$. Let $d(x,y) = 3r$ for some $r > 0$, and let $S_r(x)$ and $S_r(y)$ be open spheres centered on x and y, respectively. Then, clearly $S_r(x) \cap S_r(y) = \emptyset$ because the radius of $S_r(x)$ and $S_r(y)$ is r and the distance between x and y is $3r$.

1.4 Closed Sets and Closure of Sets

Definition 1.7. Let A be a subset of a metric space X. A point $x \in X$ is said to be a *limit point* (*accumulation point* or *cluster point*) of A if each open sphere centered on x contains at least one point of A different from x.

In other words, $x \in X$ is a limit point of A if

$$(S_r(x) - \{x\}) \cap A \neq \emptyset \quad \text{for all } r > 0.$$

The set of all limit points of A is called *derived set* and it is denoted by A'.

Example 1.20. 1. In the usual metric space $\mathbb{R}$,

(a) if $A = \left\{1, \frac{1}{2}, \frac{1}{3}, \cdots\right\}$, then $A' = \{0\}$;
(b) if $A = \mathbb{N}$ or $\mathbb{Z}$, then $A' = \emptyset$;
(c) if A is the set of all rational or irrational numbers, then $A' = \mathbb{R}$
(d) every point on the real line is a limit point, and therefore, $\mathbb{R}' = \mathbb{R}$;
(e) if A is a cantor set C, then $A' = C$.

2. If A is a subset of a discrete metric space, then A has no limit points, since every open sphere of radius 1 consists only the center. Thus $A' = \emptyset$.

Remark 1.13. By the definition of a limit point, we follow that any open sphere centered on a limit point of A must contain infinitely many points of A, that is, to say, a point $x \in X$ is a limit point of A if $S_r(x) \cap A$ is an infinite set for each $r > 0$.

Let $S_r(x)$ contain a point x_1 of A different from x. If $d(x,x_1) = r_1$, the open sphere $S_{r_1}(x)$ contains a point x_2 of A different from x and x_1. And so an indefinitely. It should be noted that a limit point of A is not necessarily a point of A. For example, in Example 1.20 (a), 0 is the only limit point of the set $A = \left\{1, \frac{1}{2}, \frac{1}{3}, \cdots\right\}$ which is not in A.

In view of the above remark, we have the following definition.

Definition 1.8. A point $x \in X$ is said to be an *isolated point* of X if there exists an open sphere centered on x contains no point of A other than x itself, that is, if $S_r(x) \cap A = \{x\}$ for some $r > 0$.

Remark 1.14. If a point $x \in X$ is not a limit point of A then it is an isolated point. Hence every point of a metric space X is either a limit point or an isolated point of X.

Example 1.21. Consider the metric space $X = \{0, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \cdots\}$ with the usual metric given by the absolute value. Then 0 is the only limit point of X while all other points are the isolated point of X.

Definition 1.9. Let A be a subset of a metric space X. The *closure* of A, denoted by $\overline{A}$ or clA, is the union of A and the set of all limit points of A, that is, $\overline{A} = A \cup A'$.

In other words, $x \in \overline{A}$ if every open sphere $S_r(x)$ with center on x and radius $r > 0$ contains a point of A, that is, $x \in \overline{A}$ if and only if $S_r(x) \cap A \neq \emptyset$ for every $r > 0$.

Remark 1.15. Let A and B be subsets of a metric space X. Then

(i) $\overline{\emptyset} = \emptyset$
(ii) $\overline{X} = X$
(iii) $\overline{(\overline{A})} = \overline{A}$
(iv) $A \subseteq B$ implies $\overline{A} \subseteq \overline{B}$
(v) $\overline{A \cup B} = \overline{A} \cup \overline{B}$
(vi) $\overline{A} = (\overline{A})'$
(vii) $\overline{A \cap B} \subseteq \overline{A} \cap \overline{B}$, but $\overline{A \cap B} \not\supseteq \overline{A} \cap \overline{B}$; For example, in the usual metric space $\mathbb{R}$, consider the sets $A = (0,1)$ and $B = (1,2)$. Then $\overline{A} \cap \overline{B} = [0,1] \cap [1,2] = \{1\}$, but $\overline{A \cap B} = \emptyset$ and hence $\overline{A \cap B} \not\supseteq \overline{A} \cap \overline{B}$.

Remark 1.16. In a metric space (X,d), $S_r(x) \subseteq S_r[x]$ for any $x \in X$ and $r > 0$. By Remark 1.15 (iv), $\overline{S_r(x)} \subseteq \overline{S_r[x]} = S_r[x]$ because $S_r[x]$ is a closed set. But in general $\overline{S_r(x)} \neq S_r[x]$. For example, let $\mathbb{R}$ be a discrete metric space. Then $S_r(0) = \{x \in \mathbb{R} : d(0,x) < 1\} = \{0\}$ which is open as well closed in X, while $S_r[0] = \{x \in \mathbb{R} : d(0,x) \leq 1\} = \mathbb{R}$. Therefore, $\overline{S_r(x)} \neq S_r[x]$.

Theorem 1.5. *Let (X,d) be a metric space and A be a subset of X. Then $x \in \overline{A}$ if and only if $\rho(x,A) = 0$.*

Proof. Let $x \in \overline{A}$. If $x \in A$, then obviously we have $d(x,A) = 0$. Assume $x \notin A$. Then x is a limit point of A. Thus for any $\varepsilon > 0$, there exists a $y \in S_\varepsilon(x) \cap A$, that is, $d(x,y) < \varepsilon$. Therefore $d(x,A) < \varepsilon$ for any $\varepsilon > 0$. Hence $d(x,A) = 0$.

Conversely, suppose that $d(x,A) = 0$. If $x \in A$, then $x \in \overline{A}$. Assume $x \notin A$. Then by the property of the infimum, for any $\varepsilon > 0$, there is a $y \in A$ such that $d(x,y) < \varepsilon$, that is, $y \in S_\varepsilon(x) \cap A$. Since $x \notin A$, then $y \neq x$. Therefore, x is a limit point of A, and thus $x \in \overline{A}$.

Definition 1.10. Let A be a subset of a metric space X. The set A is said to be *closed* if it contains all its limit points, that is, $A' \subseteq A$.

It is obvious that A is closed if and only if $\overline{A} = A$.

Example 1.22. In the usual metric space $\mathbb{R}$,

(a) the sets of all rational and irrational numbers are not closed;
(b) the set $A = \{1, \frac{1}{2}, \frac{1}{3}, \cdots, \}$ is not closed, since $A' = \{0\} \not\subseteq A$;
(c) if $A = \bigcap_{n=1}^{\infty} [-3, \frac{1}{n})$, then $\overline{A} = [-3, 0]$;
(d) if $A = (0,1) \cap \mathbb{Q}$, then $\overline{A} = [0,1]$;
(e) the cantor set C is closed since $A' = A \subseteq A$.

Remark 1.17. In a metric space X, every finite set, empty set and whole space are closed sets.

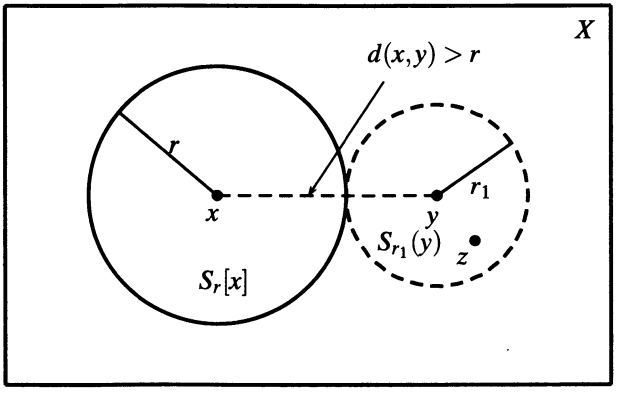
Theorem 1.6. *Let A be a subset of a metric space X. Then, A is closed if and only if the complement of A is an open set.*

Proof. Let A be closed and $x \in A^c$ be arbitrary. Then $x \notin A$ and also x cannot be a limit point of A since A is closed. Then there exists an open sphere $S_r(x)$ such that $S_r(x) \cap A = \emptyset$. This implies that $S_r(x) \subseteq A^c$ for some $r > 0$. Since $x \in A^c$ is arbitrary, each point of A^c is the center of some open sphere which is contained in A^c. Hence A^c is open.

Conversely, assume that A^c is open. Let $x \in X$ be a limit point of A. If $x \in A$, then A contains all its limit points and hence A is closed. If $x \notin A$, then $x \in A^c$. Since A^c is open, there exists an open sphere $S_r(x)$ such that $S_r(x) \subseteq A^c$. Consequently, $S_r(x) \cap A = \emptyset$ for some $r > 0$. Hence x cannot be a limit point of A which contradicts to our assumption. Therefore $x \in A$. This proves that A is closed.

Theorem 1.7. *In a metric space (X,d), every closed sphere is a closed set.*

Proof. Let $S_r[x]$ be a closed sphere in X. Then it is sufficient to show that $(S_r[x])^c$, the complement of $S_r[x]$, is an open set. Let $y_1 \in (S_r[x])^c$ be arbitrary. Then $y \notin S_r[x]$ and therefore $d(x,y) > r$.



Set $r_1 = d(x,y) - r > 0$. Let $z \in S_{r_1}(y)$. Then $d(z,y) < r_1$. By the triangle inequality

$$d(x,y) \leq d(x,z) + d(z,y)$$

and we have

$$d(x,z) \geq d(x,y) - d(z,y) > d(x,y) - r_1 = r.$$

Therefore $z \notin S_r[x]$ and hence $z \in (S_r[x])^c$. Thus $S_{r_1}(y) \subseteq (S_r[x])^c$. But $y \in (S_r[x])^c$ being arbitrary, each point of $(S_r[x])^c$ is the center of some open sphere contained in $(S_r[x])^c$. Hence $(S_r[x])^c$ is an open set.

By using De Morgan's law

$$\bigcap_{\alpha \in \Lambda} (A_\alpha^c) = \left(\bigcup_{\alpha \in \Lambda} A_\alpha \right)^c$$

and

$$\bigcup_{i=1}^{n} A_i^c = \left(\bigcap_{i=1}^{n} A_i \right)^c$$

and Theorem 1.3, we have the following result.

Theorem 1.8. *In a metric space X,*

(a) the arbitrary intersection of closed sets in X is closed; and
(b) the finite union of closed sets in X is closed.

Remark 1.18. The arbitrary union of closed sets need not be closed.

Example 1.23. Consider the family $\left\{ \left[\frac{1}{n}, 2\right] : n \in \mathbb{N} \right\}$ of closed sets in the usual metric space $\mathbb{R}$. Then

$$\bigcup \left\{ \left[\frac{1}{n}, 2\right] : n \in \mathbb{N} \right\} = (0,2]$$

which is not a closed set.

Theorem 1.9. *Let (X,d) be a metric space and A be a subset of X. Then $\overline{A}$ is the smallest closed subset of X containing A.*

Proof. We first prove that $\overline{A}$ is closed. Let x be a limit point of $\overline{A}$. Then for a given $\varepsilon > 0$, $\left(S_{\varepsilon/2}(x) - \{x\}\right) \cap \overline{A} \neq \emptyset$. This implies that there exists $y \in \overline{A}$ such that $y \in \left(S_{\varepsilon/2}(x) - \{x\}\right)$, that is, $d(x,y) < \frac{\varepsilon}{2}$. But since $y \in \overline{A}$, we have $S_{\varepsilon/2}(y) \cap A \neq \emptyset$, that is, there exists $z \in A$ such that $z \in S_{\varepsilon/2}(y)$. This implies that $d(y,z) < \frac{\varepsilon}{2}$. By triangle inequality, we have

$$\begin{aligned} d(x,z) &\leq d(x,y) + d(y,z) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

This means that, for every $\varepsilon > 0$, the open sphere $S_\varepsilon(x)$ contains a point z of A. Hence x is a limit point of A and therefore $x \in \overline{A}$. This proves that $\overline{A}$ is a closed set.

Now, we show that $\overline{A}$ is the smallest set containing A. Assume that B is any closed subset of X such that $A \subseteq B$, then it is sufficient to prove that $\overline{A} \subseteq B$. Let $x \in \overline{A}$, then either $x \in A$ or x is a limit point of A. If $x \in A$, then $x \in B$ and hence $\overline{A} \subseteq B$. If x is a limit point of A, then for a given $\varepsilon > 0$, $(S_\varepsilon(x) - \{x\}) \cap A \neq \emptyset$, that is, there exists a point $y \in A$ such that $y \in \left(S_{\varepsilon/2}(x) - \{x\}\right)$. Then $d(x,y) < \varepsilon$. But since $A \subseteq B$ and $y \in A$, we have $y \in B$. Therefore, $\rho(x,B) = \inf_{y \in B} d(x,y) = 0$ and by Theorem 1.5 $x \in \overline{B}$. Since B is a closed set, $x \in B$ and thus $\overline{A} \subseteq B$.

Definition 1.11. Let A be a subset of a metric space X. A point $x \in X$ is called a *boundary point* of A if it is neither an interior point of A nor of $X \setminus A$, that is, $x \notin A^\circ$ and $x \notin (X \setminus A)^\circ$.

In other words, $x \in X$ is a *boundary point* of A if every open sphere centered on x intersects both A and $X \setminus A$.

The set of all boundary points of A is called the *boundary of* A and it is denoted by $b(A)$.

Remark 1.19. It is clear from the definition that $b(A) = \overline{A} \cap \overline{(X \setminus A)} = \overline{A} \cap \overline{A^c}$.

Example 1.24. 1. Let $\mathbb{R}$ be the usual metric space and $A \subseteq \mathbb{R}$.

(a) If $A = [a,b]$, $[a,b)$, $(a,b]$ or (a,b), then $b(A) = \{a,b\}$.
(b) If $A = \mathbb{N}$ (respectively $\mathbb{Z}$), then $b(A) = \mathbb{N}$ (respectively, $\mathbb{Z}$).
(c) If $A = \left\{1, \frac{1}{2}, \frac{1}{3}, \cdots, \right\}$, then $b(A) = \left\{0, 1, \frac{1}{2}, \frac{1}{3}, \cdots, \right\}$.
(d) If $A = \mathbb{Q}$, then $b(A) = \mathbb{R}$.
(e) If A is a set of all irrational numbers, then $b(A) = \mathbb{R}$.

2. Let (X,d) be a discrete metric space and $A \subseteq X$. Then $b(A) = \emptyset$.

Exercise 1.36. Verify that every subset of the discrete metric space is closed.

Exercise 1.37. Prove that every singleton set and every finite set in a metric space are closed.

Exercise 1.38. Let A be a subset of a metric space X. Prove that $\overline{A}$ is the intersection of all closed subsets of X containing A.
Hint: Let $D = \cap\{B \subseteq X : B \text{ is closed and } A \subseteq B\}$. Then clearly D is the smallest closed subset of X containing A. From the proof of Theorem 1.9, $\overline{A} = D$.

Exercise 1.39. Determine the derive set of the following sets.

(a) A finite set $A = \{1,2,\cdots,n\}$.
(b) $R = \left\{(x_1,x_2) \in \mathbb{R}^2 : x_1, x_2 \text{ are rational coordinates}\right\}$.

Exercise 1.40. Let A be a subset of a metric space X. Prove that

(i) $\overline{(X \setminus A)} = X \setminus A^\circ$, that is, $\overline{(A^c)} = (A^\circ)^c$.
(ii) $(A^c)^\circ = \left(\overline{A}\right)^c$

Exercise 1.41. Let (X,d) be a metric space and A be a closed subset of X. Prove that $x \in A$ if and only if $d(x,A) = 0$ and hence

$$x \in X \setminus A \text{ if and only if } d(x,A) > 0.$$

Exercise 1.42. Let (X,d) be a metric space, $x \in X$ and $A \subseteq X$ be a nonempty set. Prove that $d(x,A) = 0$ if and only if every neighborhood of x contains a point of A.

Exercise 1.43. Let A be a subset of a metric space X. Prove that $\overline{A} = X$ if and only if $(X \setminus A)^\circ = \emptyset$, that is, $(A^c)^\circ = \emptyset$.

Exercise 1.44. Let X be a metric space. Prove that every open sphere in X is included in a closed sphere in X and that every closed sphere in X is included in an open sphere in X.

Exercise 1.45. Let A be a nonempty subset of a metric space X. Must $(A^\circ)^c$ be equal to cl $\left((\overline{A})^c\right)$?
Hint: Consider $\mathbb{Q}$ as a subset of the usual metric space $\mathbb{R}$. Then $\mathbb{Q}^\circ = \emptyset$ and so $(\mathbb{Q}^\circ)^c = \mathbb{R}$, whereas $\overline{\mathbb{Q}} = \mathbb{R}$ and $\left(\overline{\mathbb{Q}}\right)^c = \emptyset$.

Exercise 1.46. Let (X,d) be a metric space and $A \subseteq X$. Prove that

(a) $\delta(A) = \delta(\overline{A})$;
(b) $b(A) = b(X \setminus A) = \overline{A} \cap \overline{(X \setminus A)}$;
(c) $b(A) = \overline{A} \setminus A^\circ = \overline{(X \setminus A)} \setminus (X \setminus A)^\circ$;
(d) $X \setminus b(A) = A^\circ \cup (X \setminus A)^\circ$;
(e) $A = A^\circ \cup b(A)$;
(f) $\overline{A} = A \cup b(A)$;
(g) $A^\circ = A \setminus b(A)$;
(h) $A^\circ \cap b(A) = \emptyset$;
(i) A is closed if and only if $b(A) \subseteq A$;
(j) A is open if and only if $A \cap b(A) = \emptyset$.

Exercise 1.47. Prove that any closed subset A of a metric space (X,d) is a countable intersection of open sets.

Proof. For each $n \in \mathbb{N}$, let $O_n = \bigcup_{x \in A} S_{1/n}(x)$. Then O_n is open and $A \subset O_n$ for all $n \in \mathbb{N}$. Therefore, $A \subseteq \bigcap_{n \in \mathbb{N}} O_n$. We claim that $A = \bigcap_{n \in \mathbb{N}} O_n$. Assume contrary that there exists $y \in \bigcap_{n \in \mathbb{N}} O_n \setminus A$. Since A is closed, A^c is open and contains y. Hence there exists $\varepsilon > 0$ such that $S_\varepsilon(y) \subseteq A^c$. Since $\varepsilon > 0$, there exists $n \in \mathbb{N}$ such that $1/n < \varepsilon$. Since $y \in O_n$, there exists $x \in A$ such that $d(x,y) < 1/n < \varepsilon$. So $S_\varepsilon(y) \cap A \neq \emptyset$, which implies $A \cap A^c \neq \emptyset$, a contradiction.

Exercise 1.48. Let A be nonempty subset of a metric space X. Prove or answer the following statements.

(a) $b(A)$ is a closed set.
(b) $b(A) = b(X \setminus A)$.
(c) If $x \in b(A)$, does x have to be a limit point?
(d) $x \in b(A)$ if and only if for every $\varepsilon > 0$, $S_\varepsilon(x)$ contains points of A and of $X \setminus A$.

Proof. (a) Since $b(A) = \overline{A} \cap \overline{A^c}$ and the intersection of two closed sets is closed, we have $b(A)$ is a closed set.

(b) This follows from the definition of the boundary and the fact that $X \setminus (X \setminus A) = A$.

(c) No. Indeed let $A = \{0\} \subseteq \mathbb{R}$. Then A has no limit points, but $b(A) = \{0\}$.

(d) Let $x \in b(A)$. Since $b(A) = \overline{A} \cap \overline{(X \setminus A)}$, we have $x \in \overline{A}$ and $x \in \overline{(X \setminus A)}$. Then x is a limit point of A as well as of $X \setminus A$. Then for any $\varepsilon > 0$, we have $S_\varepsilon(x) \cap A \neq \emptyset$ and $S_\varepsilon(x) \cap (X \setminus A) \neq \emptyset$.

Conversely let $x \in X$ such that for any $\varepsilon > 0$, $S_\varepsilon(x)$ contains points of A and of $X \setminus A$. It is enough to prove that $x \in \overline{A}$. If $x \in A$, then we have nothing to prove. Assume $x \notin A$. Let $\varepsilon > 0$. We know that $S_\varepsilon(x) \cap A \neq \emptyset$, since $x \notin A$, then $S_\varepsilon(x) \cap A$ contains a point of A other than x. Hence x is a limit point of A, that is, $x \in \overline{A}$.

1.5 Subspaces

Let (X,d) be a metric space and Y be a subset of X. We may convert Y into a metric space by restricting the distance function d to $Y \times Y$. In this manner each subset Y of X can be made a metric space $\left(Y, d_{|_{Y\times Y}}\right)$. On the other hand, we may be given two metric spaces (X,d) and (Y,d'). If Y is a subset of X, it makes sense to ask whether or not d' is the restriction of d.

Definition 1.12. Let (X,d) be a metric space and Y be a subset of X. The *relative metric* d_Y on Y is the restriction of the metric function d on $Y \times Y$, that is,

$$d_Y(x,y) = d(x,y) \quad \text{for all } x,y \in Y.$$

It is easy to see that d_Y is a metric on Y. The space (Y,d_Y) is called the *metric subspace* of the metric space (X,d).

In other words, let (X,d) and (Y,d') be metric spaces. We say that (Y,d') is a subspace of (X,d) if

(i) Y is a subset of X;
(ii) $d' = d_{|_{Y\times Y}}$ restriction of d on $Y \times Y$.

Example 1.25. (a) Let $\mathbb{R}$ be an usual metric space. If $Y = [0,1],(0,1],[0,1)$ or $(0,1)$ and $d_Y(x,y) = |x-y| = d(x,y)$ for all $x,y \in Y$. Then (Y,d_Y) is a subspace of $(\mathbb{R},d)$.

(b) Let $\mathbb{R}$ be the usual metric space and $\mathbb{Q}$ be the set of rational numbers. Define $d_{\mathbb{Q}} : \mathbb{Q} \times \mathbb{Q} \to \mathbb{R}$ by

$$d_{\mathbb{Q}}(x,y) = |x-y| = d(x,y) \quad \text{for all } x,y \in \mathbb{Q}.$$

Then $(\mathbb{Q},d_{\mathbb{Q}})$ is a subspace of $(\mathbb{R},d)$.

(c) Let I^n (the unit n cube) be the set of all n-tuples $(x_1,x_2,\cdots,x_n)$ of real numbers such that $0 \le x_i \le 1$ for $i = 1,2,\ldots,n$. Define $d_c : I^n \times I^n \to \mathbb{R}$ by

$$d_c(x,y) = \max_{1\le i\le n} |x_i - y_i|$$

for all $x = (x_1,x_2,\ldots,x_n) \in I^n$ and $y = (y_1,y_2,\cdots,y_n) \in I^n$. Then (I^n,d_c) is a subspace of $(\mathbb{R}^n,d_\infty)$, where d_∞ is the max metric on $\mathbb{R}^n$, that is, $d_\infty(x,y) = \max\limits_{1\le i\le n} |x_i - y_i|$ for all $x,y \in \mathbb{R}^n$.

(d) Let S^n (the n-sphere) be the set of all $n+1$-tuples $(x_1,x_2,\ldots,x_{n+1})$ of real numbers such that $x_1^2 + x_2^2 + \cdots + x_{n+1}^2 = 1$. Define $d_S : S^n \times S^n \to \mathbb{R}$ by

$$d_S(x,y) = \sqrt{\sum_{i=1}^{n+1} (x_i - y_i)^2} = d_2(x,y),$$

where d_2 is a metric on $\mathbb{R}^n$ defined as $d_2(x,y) = \left(\sum\limits_{i=1}^{n+1} (x_i - y_i)^2\right)^{1/2}$ for all $x,y \in \mathbb{R}^n$.

Then (S^n,d_S) is a subspace of $\left(\mathbb{R}^{n+1},d_2\right)$.

(e) Let A be the set of all $(n+1)$-tuples $(x_1, x_2, \ldots, x_{n+1})$ of real numbers such that $x_{n+1} = 0$. Define $d_A : A \times A \to \mathbb{R}$ by

$$d_A(x,y) = \max_{1 \le i \le n} |x_i - y_i| = d_\infty(x,y),$$

for all $x = (x_1, x_2, \ldots, x_n, 0) \in A$ and $y = (y_1, y_2, \ldots, y_n, 0) \in A$, where d_∞ is the max metric on $\mathbb{R}^{n+1}$. Then (A, d_A) is a subspace of $(\mathbb{R}^{n+1}, d_\infty)$.

(f) Let $P[a,b]$ be the set of all polynomials defined on $[a,b]$. Define $d_P : P[a,b] \times P[a,b] \to \mathbb{R}$ by

$$d_P(f,g) = \max_{t \in [a,b]} |f(t) - g(t)| = d_\infty(f,g),$$

where d_∞ is the max metric on $C[a,b]$. Then $(P[a,b], d_P)$ is a subspace $(C[a,b], d_\infty)$. But $(P[a,b], d_P)$ is not a subspace of $(C[a,b], d)$, where $d(f,g) = \int_b^a |f(t) - g(t)|\, dt$.

Remark 1.20. If Y is a subspace of a metric space X, then a set which is open (respectively, closed) in Y may not be open (respectively, closed) in X. For example, if $Y = [0,1]$ then the set $[0, \frac{1}{2})$ is open in Y but not in $\mathbb{R}$ with the usual metric. Also, if $Y = (0,1)$ then the set $(0, \frac{1}{2}]$ is closed in Y but not in $\mathbb{R}$ with the usual metric.

Lemma 1.1. *Let (Y, d_Y) be a subspace of a metric space (X,d). If $a \in Y$ and $r > 0$, then*

$$S'_r(a) = Y \cap S_r(a),$$

where $S_r(a)$ and $S'_r(a)$ are open spheres in (X,d) and (Y, d_Y), respectively.

Proof. We have

$$\begin{aligned} S_r(a) \cap Y &= \{x \in X : d(x,a) < r\} \cap Y \\ &= \{x \in Y : d(x,a) < r\} \quad \text{because } Y \subseteq X \\ &= S'_r(a). \end{aligned}$$

Theorem 1.10. *Let (Y, d_Y) be a subspace of a metric space (X,d) and A a subset of Y. Then*

(a) A is open in Y if and only if there exists an open set G in X such that $A = G \cap Y$;
(b) A is closed in Y if and only if there exists a closed set F in X such that $A = F \cap Y$.

Proof. (a) Let $S_r(x)$ and $S'_r(x)$ be the same as in Lemma 1.1. Suppose that $A = G \cap Y$, where G is open in X. Let $x \in A$ be arbitrary. Then we have to show that x is an interior point of A, that is, $x \in A^\circ$ with respect to d_Y metric.

Since $A = G \cap Y$ and $x \in A$, we have $x \in G$ and $x \in Y$. Since G is open in X, there exists $r > 0$ such that $S_r(x) \subseteq G$. Hence, by Lemma 1.1, we have

$$S'_r(x) = S_r(x) \cap Y \subseteq G \cap Y = A.$$

It follows that x is an interior point of A as a subset of the metric space (Y, d_Y). Hence $x \in A^\circ$ with respect to d_Y metric and hence A is open in Y.

Conversely, assume that A is an open set in Y and let $x \in A$ be arbitrary. Then there exists an open sphere $S'_{r_x}(x)$ such that $S'_{r_x}(x) \subseteq A$. Now, we have

$$A = \bigcup_{x\in A} S'_{r_x}(x) = \bigcup_{x\in A} (S_{r_x}(x)\cap Y) = \left(\bigcup_{x\in A} S_{r_x}(x)\right)\cap Y$$
$$= G\cap Y, \text{ where } G = \bigcup_{x\in A} S'_{r_x}(x).$$

But G being an arbitrary union of open spheres in X is an open set in X. Hence $A = G\cap Y$, where G is an open set in X.

(b) A is closed in $Y \Leftrightarrow Y\setminus A$ is open in Y

$$\Leftrightarrow Y\setminus A = G\cap Y, \quad \text{(by part (i)) where } G \text{ is open in } X$$
$$\Leftrightarrow A = Y\setminus(G\cap Y)$$
$$\Leftrightarrow A = (X\cap Y)\setminus(G\cap Y)$$
$$\Leftrightarrow A = (X\setminus G)\cap Y$$
$$\Leftrightarrow A = F\cap Y, \text{ where } F = X\setminus G \text{ is a closed set in } X.$$

Corollary 1.1. *Let (Y,d_Y) be a subspace of a metric space (X,d) and A be a subset of X.*

(a) If A is open in Y and Y is open in X then A is open in X;
(b) If A is closed in Y and Y is closed in X then A is closed in X.

Theorem 1.11. *Let (Y,d_Y) be a subspace of a metric space (X,d) and A be a subset of Y. Then*

(a) $x\in Y$ is a limit point of A in Y if and only if x is a limit point of A in X;
(b) the closure of A in Y, denoted by $\mathrm{cl_A(Y)}$, is $\mathrm{cl_X(A)}\cap \mathrm{Y}$, where $\mathrm{cl_X(A)}$ is the closure of A in X. In other words, $\mathrm{cl_Y(A)} = \mathrm{cl_X(A)}\cap \mathrm{Y}$.

Proof. (a) Let $x\in Y$ be a limit point of A in Y. Then for every open sphere $S'_r(x)$ in Y, we have $(S'_r(x)-\{x\})\cap A \neq \emptyset$. For any given $r>0$, we have

$$(S_r(x)-\{x\})\cap A = (S'_r(x)\cap Y - \{x\})\cap A \quad (\text{since } A\subseteq Y)$$
$$= (S'_r(x)-\{x\})\cap A \neq \emptyset.$$

It follows that x is a limit point of A in X.

The converse can be established by retracting the above steps.

(b) Since $\mathrm{cl_X(A)}$ is closed in X, by Theorem 1.10, $\mathrm{cl_X(A)}\cap \mathrm{Y}$ is closed in Y. Since $\mathrm{cl_X(A)}\cap \mathrm{Y}$ contains A and since $\mathrm{cl_Y(A)}$ is the intersection of all closed subsets of Y containing A, we must have

$$\mathrm{cl_Y(A)} \subseteq \mathrm{cl_X(A)}\cap \mathrm{Y}.$$

On the other hand, $\mathrm{cl_Y(A)}$ is closed in Y, then $\mathrm{cl_Y(A)} = \mathrm{F}\cap \mathrm{Y}$, where F is a closed set in X. Since $A\subseteq \mathrm{cl_Y(A)}$, F is a closed set in X containing A. Since $\mathrm{cl_Y(A)}$ is the intersection of all closed sets containing A, we have

$$\mathrm{cl_Y(A)} \subseteq \mathrm{F}.$$

Hence $\mathrm{cl_Y(A)}\cap \mathrm{Y} \subseteq \mathrm{F}\cap \mathrm{Y} = \mathrm{cl_Y(A)}$.

Exercise 1.49. Let (Y, d_Y) be a subspace of a metric space (X, d). Then prove that a subset M of Y is a neighborhood of a point $y \in Y$ if and only if there is a neighborhood N of y in (X, d) such that $M = Y \cap N$.

Proof. Let N be a neighborhood of a point $y \in Y$ in (X, d) such that $M = Y \cap N$. Then there exists an open sphere $S_r(y)$ such that $S_r(y) \subseteq N$. Since $S'_r(y) = Y \cap S_r(y)$, we have $S'_r(y) \subseteq Y \cap N = M$. Hence M is a neighborhood of $y \in Y$ in (Y, d_Y).

Conversely, suppose that M is a neighborhood of y in (Y, d_Y). Then there exists an open sphere $S'_r(y) \subseteq M$. Let $N = M \cup S_r(y)$. Then

$$\begin{aligned} Y \cap N &= Y \cap (M \cup S_r(y)) = (Y \cap M) \cup (Y \cap S_r(y)) \\ &= M \cup S'_r(y) \text{ since } M \subseteq Y \\ &= M \text{ because } S'_r(y) \subseteq M. \end{aligned}$$

Since $S_r(y) \subseteq N$, N is a neighborhood of y in (X, d).

Chapter 2
Complete Metric Spaces

The concept of a convergent sequence plays an important role to investigate the closedness of a set, the continuity of a function and several other properties. In this chapter, we give an introduction to the convergence of sequences in arbitrary metric spaces. We investigate the properties under which a sequence is convergent. In real analysis, Cauchy criteria provides the necessary and sufficient conditions for a sequence to be convergent, that is, a sequence is convergent if and only if it is Cauchy. This Cauchy criteria does not hold for any arbitrary metric space. The metric spaces in which this Cauchy criteria holds are called complete. Such metric spaces are also considered and studied. A very important property of a complete metric space is known as Cantor's intersection theorem. Such theorem and its converse version are presented. We also consider a metric space which contains an incomplete metric space and some other points so that every Cauchy sequence is convergent in this larger space. Such spaces are called completion of an incomplete metric space.

2.1 Convergent Sequences

Definition 2.1. A *sequence* s in a set X is a mapping from the set of all natural numbers $\mathbb{N}$ into X. The image under a sequence s of a natural number n is denoted by x_n and is referred as nth term of the sequence s.

Let $\{x_n\}_{n\in\mathbb{N}}$ be a given sequence and $\{n_k\}_{k\in\mathbb{N}}$ be a sequence of positive integers such that $n_1 < n_2 < \cdots$. Then the sequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ is called a *subsequence* of the sequence $\{x_n\}$.

Definition 2.2. Let $\{x_n\}$ be a sequence in a set X. If there exists a positive integer N such that $x_n = x$ for all $n > N$, then the sequence $\{x_n\}$ is called *eventually constant*.

If $x_n = n$ for all $n \in \mathbb{N}$, then the sequence $\{x_n\}$ is called a *constant sequence*.

Obviously, a constant sequence is a special case of an eventually constant sequence.

Definition 2.3. Let (X,d) be a metric space. A sequence $\{x_n\}$ of points of X is said to be *convergent* if there is a point $x \in X$ such that for each $\varepsilon > 0$, there exists a positive integer N such that

$$d(x_n,x) < \varepsilon \quad \text{for all } n > N.$$

The point $x \in X$ is called a *limit point* of the sequence $\{x_n\}$.

A sequence which is not convergent is said to be *divergent*.

More preciously, a sequence $\{x_n\}$ in a metric space X converges to a point $x \in X$ if the sequence $\{d(x_n,x)\}$ of real numbers converges to 0 as $n \to \infty$.

Since $d(x_n, x) < \varepsilon$ is equivalent to $x_n \in S_\varepsilon(x)$, the definition of convergent sequence can be restated as follows:

A sequence $\{x_n\}$ in a metric space X *converges to a point* $x \in X$ if and only if for each $\varepsilon > 0$, there exists a positive integer N such that

$$x_n \in S_\varepsilon(x) \quad \text{for all } n > N.$$

We use the following symbols to write a convergent sequence

$$x_n \to x \qquad \text{or} \qquad \lim_{n\to\infty} x_n = x$$

and we express it by saying that x_n approaches x or that x_n converges to x.

Remark 2.1. (i) An eventually constant sequence, and hence a constant sequence, is convergent.

(ii) In a discrete metric space, a sequence can converge to a point only if it is an eventually constant sequence.

Remark 2.2. The convergence of a sequence in a metric space (X, d) depends on the space X as well as on the metric d.

Example 2.1. (a) Let $\left\{x_n = \frac{1}{n} : n \in \mathbb{N}\right\}$ be a sequence in the usual metric space $\mathbb{R}$. Then $x_n \to 0$ as $n \to \infty$. However, if we consider the metric space $X = (0,1)$ with the usual metric, then the sequence $\left\{x_n = \frac{1}{n} : n \in \mathbb{N}\right\}$ converges to $0 \notin X$. Thus, the sequence $\{x_n\}$ is not convergent in this case.

(b) Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence in the space $C[0,1]$, where

$$f_n(t) = e^{-nt} \quad \text{for all } n \in \mathbb{N}.$$

Then $f_n \to 0$ with respect to the metric d_1 on $C[0,1]$ as

$$d_1(f_n, 0) = \int_0^1 e^{-nt} dt = \frac{1}{n}\left(1 - e^{-n}\right) \to 0 \text{ as } n \to \infty.$$

On the other hand, the same sequence $\{f_n\}_{n\in\mathbb{N}}$ is not convergent with respect to the metric d_∞ on $C[0,1]$ as

$$d_\infty(f_n, 0) = \max_{t\in[0,1]} \left|e^{-nt}\right| = 1 \quad \text{for all } n \in \mathbb{N}.$$

Thus $d_\infty(f_n, 0) \not\to 0$ as $n \to \infty$.

Theorem 2.1. *A sequence in a metric space cannot converge to more than one limit point. In other words, in a metric space, every convergent sequence has a unique limit.*

Proof. Let (X, d) be a metric space and $\{x_n\}$ be a convergent sequence in X. Suppose to the contrary that $\{x_n\}$ converges to two distinct points x and y. Then, for each $\varepsilon > 0$, there exist positive integers N_1 and N_2 such that

$$d(x_n, x) < \frac{\varepsilon}{2} \quad \text{for all } n > N_1$$

and

$$d(x_n, y) < \frac{\varepsilon}{2} \quad \text{for all } n > N_2.$$

By setting $N = \max\{N_1, N_2\}$ and using the triangle inequality, we have

$$\begin{aligned} d(x,y) &\leq d(x_n,x) + d(x_n,y) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \quad \text{for all } n > N. \end{aligned}$$

It follows that $x = y$. Hence the limit is unique.

Definition 2.4. A sequence in a metric space is said to be *bounded* if the range set of the sequence is bounded.

Theorem 2.2. *In a metric space, every convergent sequence is bounded.*

Proof. Let (X,d) be a metric space and $\{x_n\}$ be a sequence in X such that $x_n \to x \in X$ as $n \to \infty$. Then there exists a positive integer N such that

$$d(x_n, x) < \frac{1}{2} \quad \text{for all } n > N.$$

Set $r = \max\left\{\frac{1}{2}, d(x_n,x) \text{ for all } 1 \leq n \leq N\right\}$. Then

$$d(x_n, x) \leq r \quad \text{for all } n \in \mathbb{N}.$$

By the triangle inequality,

$$\begin{aligned} d(x_n, x_m) &\leq d(x_n,x) + d(x,x_m) \\ &\leq 2r \quad \text{for all } n,m \in \mathbb{N}. \end{aligned}$$

Hence the diameter of the range set of the sequence $\{x_n\}$ is bounded by $2r$. Therefore, the sequence $\{x_n\}$ is bounded.

Theorem 2.3. *Let (X,d) be a metric space and A be a subset of X.*

(a) A point $x \in X$ is a limit point of A if and only if there exists a sequence $\{x_n\}$ of points of A, none of which equals x, such that $\{x_n\}$ converges to x.
(b) The set A is closed if and only if every convergent sequence of points of A has its limit in A.

Proof. (a) Let $x \in X$ be a limit point of A. Construct a sequence $\{x_n\}$ by recursion as follows:

Since $x \in X$ is a limit point of A, we have $(S_1(x) - \{x\}) \cap A \neq \emptyset$. So, we can choose $x_1 \in (S_1(x) - \{x\}) \cap A$. Likewise the points $x_1, x_2, \ldots, x_n$ can be chosen such that

$$x_i \in \left(S_{\frac{1}{i}}(x) - \{x\}\right) \cap A \quad \text{for } i = 1,2,\ldots,n.$$

Still $\left(S_{\frac{1}{n+1}}(x) - \{x\}\right) \cap A \neq \emptyset$, we can always choose $x_{n+1} \in \left(S_{\frac{1}{n+1}}(x) - \{x\}\right) \cap A$. By repeating this process infinitely many times, we construct the sequence $\{x_n\}$ by recursion such that all the points of $\{x_n\}$ are in A and $x_n \neq x$ for all n.

Let $\varepsilon > 0$ be given and N be a positive integer such that $N > \frac{1}{\varepsilon}$. Then

$$x_n \in S_{\frac{1}{n}}(x) \subset S_\varepsilon(x) \quad \text{for all } n > N.$$

Hence $\{x_n\}$ converge to x.

Conversely, assume that there is a sequence $\{x_n\}$ of point of A, none of which equals x, such that $\{x_n\}$ converges to x. Then for every $\varepsilon > 0$, there exists a positive integer N such that

$$x_n \in S_\varepsilon(x) \quad \text{for all } n > N.$$

Therefore, $(S_\varepsilon(x) - \{x\}) \cap A \neq \emptyset$ which implies that x is a limit point of A.

(b) Suppose that A is closed and $\{x_n\}$ is a sequence of points of A which converges to a point x in X. Then we have to show that $x \in A$.

If the range set of the sequence $\{x_n\}$ is infinite, then it follows from part (i) that x is a limit point of this set. Since A is closed, we have $x \in A$.

On the other hand, if the range set of the sequence $\{x_n\}$ is finite, then $x_n = x$ for all $n \geq N$, since $\{x_n\}$ is a convergent sequence. Since each term of the sequence belongs to A, so is x.

Conversely, assume that every convergent sequence of points of A converges to a point of A. We show that A is closed by showing that it contains all its limits points.

Let x be a limit point of A. Then by part (i), there is a sequence $\{x_n\}$ of points of A, none of which equals x, such that $x_n \to x$. By hypothesis $x \in A$. Hence A is closed.

Exercise 2.1. Show that the limit of a convergent sequence of distinct points in a metric space is a limit of the range set of the sequence.

Proof. Let $\{x_n\}$ be a sequence in a metric space such that $x_n \to x$ and let A be the range set of the sequence $\{x_n\}$. Then we have to show that x is a limit point of A.

Suppose that x is not a limit point of A. Then there exists an open sphere $S_\varepsilon(x)$ such that

$$(S_\varepsilon(x) - \{x\}) \cap A = \emptyset,$$

that is, $S_\varepsilon(x)$ contains no point of A other than x. Since x is a limit point of the sequence, for each $\varepsilon > 0$, there exists a positive integer N such that

$$d(x_n, x) < \varepsilon \qquad \text{or} \qquad x_n \in S_\varepsilon(x) \quad \text{for all } n > N$$

which is a contradiction.

Exercise 2.2. Let (X, d) be a metric space. If $\{x_n\}$ and $\{y_n\}$ be sequences in X such that $x_n \to x$ and $y_n \to y$, then prove that $d(x_n, y_n) \to d(x, y)$.

Proof. Since $x_n \to x$ and $y_n \to y$, for each $\varepsilon > 0$, there exist positive integers N_1 and N_2 such that

$$d(x_n, x) < \frac{\varepsilon}{2} \quad \text{for all } n > N_1$$

and

$$d(x_m, x) < \frac{\varepsilon}{2} \quad \text{for all } m > N_2.$$

Let $N = \max\{N_1, N_2\}$. Then, for all $n, m > N$,

$$\begin{aligned}|d(x_n,y_n)-d(x,y)| &\le |d(x_n,y_n)-d(x_n,y)|+|d(x_n,y)-d(x,y)|\\ &\le d(y_n,y)+d(x_n,x)\\ &< \frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.\end{aligned}$$

Hence $d(x_n,y_n)\to d(x,y)$.

2.2 Cauchy Sequences

Definition 2.5. Let (X,d) be a metric space. A sequence $\{x_n\}$ in X is said to be a *Cauchy sequence* if for each $\varepsilon>0$, there exists a positive integer N such that

$$d(x_n,x_m)<\varepsilon \quad \text{for all } n,m>N.$$

Example 2.2. Let $X=C[0,1]$ be a metric space with the metric d_∞ defined by

$$d_\infty(f,g)=\sup_{t\in[0,1]}|f(t)-g(t)| \quad \text{for all } f,g\in X.$$

The sequence $\{f_n\}$ in X given by

$$f_n(t)=\frac{nt}{n+t} \quad \text{for all } t\in[0,1],$$

is a Cauchy sequence in X. Indeed, for $m\ge n$, the function

$$t\mapsto \frac{mt}{m+t}-\frac{nt}{n+t}=\frac{(m-n)t^2}{(m+t)(n+t)}$$

being continuous on $[0,1]$, assumes its maximum at some point $t_0\in[0,1]$. So,

$$\begin{aligned}d_\infty(f_m,f_n) &= \sup_{t\in[0,1]}|f_m(t)-f_n(t)|\\ &= \frac{(m-n)t_0^2}{(m+t_0)(n+t_0)}\\ &\le \frac{t_0^2}{(n+t_0)}\le\frac{1}{n}\to 0 \quad \text{for large } n \text{ and } m.\end{aligned}$$

Moreover, the sequence $\{f_n\}$ converges to some limit. Indeed, let $f(t)=t$. Then

$$|f_n(t)-f(t)|=\left|\frac{nt}{n+t}-t\right|=\frac{t^2}{n+t}\le\frac{1}{n}\to 0 \text{ as } n\to\infty.$$

Therefore, $\{f_n\}$ converges to the limit f, where $f(t)=t$ for all $t\in[0,1]$.

Theorem 2.4. *Every convergent sequence in a metric space is a Cauchy sequence.*

Proof. Let (X,d) be a metric space and $\{x_n\}$ be a sequence in X such that $x_n\to x$ as $n\to\infty$. Then for each $\varepsilon>0$, there exists a positive integer N such that

$$d(x_n,x) < \frac{\varepsilon}{2} \quad \text{for all } n > N.$$

By the triangle inequality,

$$\begin{aligned} d(x_m,x_n) &\leq d(x_m,x) + d(x_n,x) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \quad \text{for all } n,m > N. \end{aligned}$$

Hence $\{x_n\}$ is a Cauchy Sequence.

Remark 2.3. Every Cauchy sequence in a metric space need not be convergent.

Example 2.3. (a) Consider the sequence $\{x_n\}$ in the usual metric space $\mathbb{Q}$, where

$$\begin{aligned} x_1 &= 1.4 \\ x_2 &= 1.41 \\ x_3 &= 1.414 \\ x_4 &= 1.4142 \\ x_5 &= 1.41421 \\ \cdots &\quad \cdots\cdots\cdots\cdots\cdots \end{aligned}$$

Then the sequence $\{x_n\}$ converges to $\sqrt{2}$. Hence $\{x_n\}$ is a Cauchy sequence. However, it does not converge to a point of $\mathbb{Q}$.

(b) Let $X = (0,1]$ be a metric space with the usual metric and $\left\{x_n = \frac{1}{n} : n \in \mathbb{N}\right\}$ be a sequence in X. Then $\{x_n\}$ is a Cauchy sequence since for each $\varepsilon > 0$, we have

$$d(x_m,x_n) = \left|\frac{1}{m} - \frac{1}{n}\right| < \varepsilon \quad \text{for all } m,n > \frac{1}{\varepsilon}.$$

On other hand, $x_n \to 0 \notin X$.

Remark 2.4. In Example 2.3 (b), if we consider $X = [0,1]$, then the sequence $\left\{\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \cdots\right\}$ is Cauchy as well as convergent.

Theorem 2.5. *Let (X,d) be a metric space and $\{x_n\}$ be a convergent sequence in X such that $x_n \to x$ as $n \to \infty$. If $\{x_{n_k}\}$ is any subsequence of $\{x_n\}$, then $x_{n_k} \to x$ as $k \to \infty$.*

Proof. Since $x_n \to x$ as $n \to \infty$, for each $\varepsilon > 0$, there exists a positive integer N such that $d(x_n,x) < \frac{\varepsilon}{2}$ for all $n > N$. Also, since every convergent sequence is Cauchy, we have

$$\begin{aligned} d(x_{n_k},x) &\leq d(x_{n_k},x_n) + d(x_n,x) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \quad \text{for all } n,n_k > N. \end{aligned}$$

Hence $x_{n_k} \to x$ as $k \to \infty$.

Remark 2.5. Theorem 2.5 says that every subsequence of a convergent sequence is convergent. However, if a subsequence of a sequence in a metric space (X,d) is convergent, then the sequence itself need not be convergent.

Example 2.4. Consider the sequence $\{x_n = (-1)^n : n \in \mathbb{N}\}$ in $\mathbb{R}$ with the usual metric. Let $\{x_{2n}\}$ be a subsequence of the sequence $\{x_n\}$ given by

$$x_{2n} = 1 \quad \text{for all } n.$$

Then $x_{2n} \to 1$ as $n \to \infty$. However, $\{x_n\}$ is not a convergent sequence.

Theorem 2.6. *Let $\{x_n\}$ be a Cauchy sequence in a metric space (X,d). Then the sequence $\{x_n\}$ is convergent if and only if it has a convergent subsequence.*

Proof. Let $\{x_{n_k}\}$ be a convergent subsequence of the sequence $\{x_n\}$. Suppose that $x_{n_k} \to x$ as $k \to \infty$. Then for each $\varepsilon > 0$, there exists a positive integer N such that

$$d(x_{n_k}, x) < \frac{\varepsilon}{2} \quad \text{for all } n_k > N.$$

Since $\{x_n\}$ is a Cauchy sequence, we have

$$d(x_{n_k}, x_n) < \frac{\varepsilon}{2} \quad \text{for all } n, n_k > N.$$

By the triangle inequality, we have

$$\begin{aligned} d(x_n, x) &\leq d(x_n, x_{n_k}) + d(x_{n_k}, x) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \quad \text{for all } n > N. \end{aligned}$$

Hence the sequence $\{x_n\}$ is convergent.

The converse part follows from Theorem 2.5.

Exercise 2.3. Consider the sequence $\{x_n\}$ in the usual metric space $\mathbb{Q}$, where

$$x_1 = 0.1,\ x_2 = 0.101,\ x_3 = 0.101001,\ x_4 = 0.1010010001, \ldots.$$

Then prove that the sequence $\{x_n\}$ is Cauchy which does not converge to a point of $\mathbb{Q}$.

Exercise 2.4. Prove that every Cauchy sequence in a discrete metric space is convergent.

Proof. Let (X,d) be a discrete metric space and $\{x_n\}$ be a Cauchy sequence in X. Recall that the discrete metric d is defined by

$$d(x,y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y. \end{cases}$$

Let $\varepsilon = \frac{1}{2}$. Since $\{x_n\}$ is a Cauchy sequence, there exists a positive integer N such that

$$d(x_n, x_m) < \frac{1}{2} \quad \text{for all } n, m > N.$$

From the definition of d, we have $x_n = x_m$ for all $n, m > N$. In other words, $\{x_n\}$ is of the form $\{x_1, x_2, \ldots, x_N, x, x, \ldots\}$. Hence $x_n \to x$ as $n \to \infty$.

Exercise 2.5. Prove that every Cauchy sequence in a metric space is bounded.

Exercise 2.6. Let d and d^* be two metrics on the same underlying set X and there exist two real numbers $K_1, K_2 > 0$ such that

$$K_1 d(x,y) \leq d^*(x,y) \leq K_2 d(x,y) \quad \text{for all } x, y \in X.$$

Prove that the sequence $\{x_n\}$ is Cauchy (respectively, convergent) in (X,d) if and only if it is Cauchy (respectively, convergent) in (X,d^*).

Exercise 2.7. Let $\{x_n\}$ be a Cauchy sequence in a metric space (X,d) and $\{x_{n_k}\}$ be a subsequence of $\{x_n\}$. Show that $\lim_{n\to\infty} d(x_n, x_{n_k}) = 0$.

Proof. Let $\varepsilon > 0$. Since $\{x_n\}$ is a Cauchy sequence in X, there exists a positive integer N such that

$$d(x_n, x_m) < \varepsilon \quad \text{for all } n, m > N - 1.$$

Now $n_N \geq N > N - 1$ and therefore

$$d(x_N, x_{n_N}) < \varepsilon.$$

In other words $\lim_{n\to\infty} d(x_n, x_{n_k}) = 0$.

Exercise 2.8. Let $\{x_n\}$ and $\{y_n\}$ be sequences in a metric space (X,d) such that $\{y_n\}$ is Cauchy and $d(x_n, y_n) \to 0$ as $n \to \infty$. Prove that

(a) $\{x_n\}$ is a Cauchy sequence in X;
(b) $\{x_n\}$ converges to $x \in X$ if and only if $\{y_n\}$ converges to x.

Proof. (a) Let $\varepsilon > 0$. Since $\{y_n\}$ is a Cauchy sequence, there exists a positive integer N_1 such that

$$d(y_m, y_n) < \frac{\varepsilon}{3} \quad \text{for all } m, n > N_1.$$

Since $d(x_n, y_n) \to 0$ as $n \to \infty$, there exists a positive integer N_2 such that $\frac{1}{N_2} < \frac{\varepsilon}{3}$ and

$$d(x_n, y_n) < \frac{\varepsilon}{3} \quad \text{for all } n > N_2.$$

By the triangle inequality, we have

$$d(x_m, x_n) \leq d(x_m, y_m) + d(y_m, y_n) + d(y_n, x_n).$$

Hence for all $n, m > N_2$, we have

$$d(x_m, x_n) < \frac{\varepsilon}{3} + d(y_m, y_n) + \frac{\varepsilon}{3}.$$

Let $N_0 = \max\{N_1, N_2\}$. Then for all $n, m > N_0$, we have

$$d(x_m, x_n) < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

Thus $\{x_n\}$ is a Cauchy sequence.

(b) By the triangle inequality, we have

$$d(y_n, x) \leq d(y_n, x_n) + d(x_n, x)$$

and hence

$$\lim_{n\to\infty} d(y_n, x) \leq \lim_{n\to\infty} d(y_n, x_n) + \lim_{n\to\infty} d(x_n, x).$$

But $\lim_{n\to\infty} d(y_n, x_n) = 0$ and if $\lim_{n\to\infty} d(x_n, x) = 0$, we have $\lim_{n\to\infty} d(y_n, x) = 0$. Thus, $y_n \to x$ as $n \to \infty$.

Exercise 2.9. Let $\{x_n\}$ and $\{y_n\}$ be Cauchy sequences in a metric space (X, d). Prove that $\{d(x_n, y_n)\}$ is a Cauchy sequence.

Proof. From Exercise 1.1, for any $n, m \in \mathbb{N}$ we have

$$|d(x_n, y_n) - d(x_m, y_m)| \leq d(x_n, x_m) + d(y_n, y_m).$$

Since $\{x_n\}$ and $\{y_n\}$ are Cauchy sequences, for any given $\varepsilon > 0$, there exist positive integers N_1 and N_2 such that for any $n, m \geq N_1$ we have

$$d(x_n, x_m) < \frac{\varepsilon}{2},$$

and for any $n, m \geq N_2$ we have

$$d(y_n, y_m) < \frac{\varepsilon}{2}.$$

Set $N = \max\{N_1, N_2\}$. Then for any $n, m \geq N$, we have

$$|d(x_n, y_n) - d(x_m, y_m)| \leq d(x_n, x_m) + d(y_n, y_m) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

This implies that $\{d(x_n, y_n)\}$ is a Cauchy sequence in $\mathbb{R}$.

Exercise 2.10. Let (X, d) be a metric space and d^* be another metric on X defined by

$$d^*(x, y) = \min\{1, d(x, y)\}.$$

Show that $\{x_n\}$ is a Cauchy sequence in (X, d) if and only if it is a Cauchy sequence in (X, d^*).

Exercise 2.11. Let $X = C[0, 1]$ be a metric space with the following metric

$$d_2(f, g) = \left[\int_0^1 |f(t) - g(t)|^2\right]^{1/2} \quad \text{for all } f, g \in X \text{ and for all } t \in [0, 1].$$

Show that the sequence $\{f_n\}$ defined by

$$f_n(t) = \begin{cases} 1-nt & \text{if } 0 \le t \le \frac{1}{n} \\ 0 & \text{if } \frac{1}{n} \le t \le 1, \end{cases}$$

is Cauchy with respect to the metric d_2 but not Cauchy with respect to the metric d_∞. *Hint:*

$$d_2(f_n, f_m) = \left[\int_0^{\frac{1}{n}} (n-m)^2 t^2 dt + \int_{\frac{1}{n}}^{\frac{1}{m}} (1-mt)^2 dt\right]^2$$

$$= \left[\frac{1}{3}\left(\frac{1}{m} - \frac{2}{n} + \frac{m}{n^2}\right)\right]^{\frac{1}{2}}$$

and

$$d_\infty(f_n, f_m) = \sup_{0 \le t \le 1} |f_n(t) - f_m(t)| = 1 - \frac{m}{n}.$$

2.3 Complete Metric Spaces

Definition 2.6. A metric space (X,d) is said to be *complete* if every Cauchy sequence in X converges to a point in X.

Remark 2.6. In view of Theorem 2.6, a metric space (X,d) is complete if and only if every Cauchy sequence in X has a convergent subsequence.

Example 2.5. (a) The usual metric spaces $\mathbb{R}$ and $\mathbb{C}$ are complete.

(b) The set of integer $\mathbb{Z}$ with the usual metric is a complete metric space.

Verification. Let $\{x_n\}$ be a Cauchy sequence of integers, that is, each term of the sequence belongs to $\mathbb{Z} = \{\ldots,-2,-1,0,1,2,\ldots\}$. Then the sequence must be of the form $\{x_1, x_2, x_3, \ldots, x_n, x, x, x, \ldots\}$. Indeed, if we choose $\varepsilon = \frac{1}{2}$, then

$$x_n, x_m \in \mathbb{Z} \text{ and } |x_n - x_m| < \frac{1}{2} \quad \text{implies} \quad x_n = x_m.$$

Hence the sequence $\{x_1, x_2, \ldots, x_n, x, x, x, \ldots\}$ converges to x.

(c) The Euclidean space $\mathbb{R}^n$ with the usual metric

$$d(x,y) = \left(\sum_{k=1}^{n} (x_k - y_k)^2\right)^{\frac{1}{2}}$$

for all $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ in $\mathbb{R}^n$, is a complete metric space.

Verification. Let $\left\{x^{(m)}\right\}$ be a Cauchy sequence in $\mathbb{R}^n$, where $x^{(m)} = \left(x_1^{(m)}, x_2^{(m)}, \ldots, x_n^{(m)}\right)$, that is, $x^{(1)} = \left(x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)}\right)$, $x^{(2)} = \left(x_1^{(2)}, x_2^{(2)}, \ldots, x_n^{(2)}\right)$ and so on. Then for each $\varepsilon > 0$, there exists a positive integer N such that

$$d\left(x^{(m)},x^{(p)}\right)=\left[\sum_{k=1}^{n}\left(x_k^{(m)}-x_k^{(p)}\right)^2\right]^{\frac{1}{2}}<\varepsilon \quad \text{for all } m,p>N.$$

On squaring both the sides, we get

$$\sum_{k=1}^{n}\left(x_k^{(m)}-x_k^{(p)}\right)^2<\varepsilon^2\Rightarrow\left(x_k^{(m)}-x_k^{(p)}\right)^2<\varepsilon^2\Rightarrow\left|x_k^{(m)}-x_k^{(p)}\right|<\varepsilon$$

for all $m,p>N$ and all $k=1,2,\ldots,n$.

It follows that for each fixed k $(1\leq k\leq n)$, the sequence $\left\{x_k^{(m)}\right\}_{m\in\mathbb{N}}$ is a Cauchy sequence in the usual metric space $\mathbb{R}$. Since $\mathbb{R}$ is complete, it converges to some point in $\mathbb{R}$. Let $x_k^{(m)}\to x_k$ as $m\to\infty$ for each $k=1,2,\ldots,n$, that is, $x_1^{(m)}\to x_1, x_2^{(m)}\to x_2,\ldots,x_n^{(m)}\to x_n$. Then for each $k=1,2,\ldots,n$, $\left|x_k^{(m)}-x_k\right|\to 0$ as $m\to\infty$. Let $x=(x_1,x_2,\ldots,x_n)$, then clearly $x\in\mathbb{R}^n$ and hence

$$d\left(x^{(m)},x\right)=\left[\sum_{k=1}^{n}\left(x_k^{(m)}-x_k\right)^2\right]^{\frac{1}{2}}\to 0 \quad \text{as } m\to\infty.$$

Therefor, $x^{(m)}\to x$ as $m\to\infty$, and thus the sequence $\left\{x^{(m)}\right\}$ is convergent. Hence, $\mathbb{R}^n$ is complete.

(d) The unitary space $\mathbb{C}^n$ is a complete metric space.

Example 2.6. By Exercise 2.4, every Cauchy sequence is convergent in a discrete metric space and hence every discrete metric space is complete.

Example 2.7. The space ℓ^p $(1\leq p<\infty)$ of all sequences $\{x_n\}$ of real or complex numbers such that $\sum_{k=1}^{\infty}|x_k|^p<\infty$ with the metric

$$d_p(x,y)=\left(\sum_{k=1}^{\infty}|x_k-y_k|^p\right)^{\frac{1}{p}} \quad \text{for all } x=\{x_n\},y=\{y_n\}\in\ell^p$$

is a complete metric space.

Verification. Let $\left\{x^{(m)}\right\}_{m\in\mathbb{N}}$ be a Cauchy sequence in ℓ^p, where $x^{(m)}=\left\{x_1^{(m)},x_2^{(m)},\ldots\right\}=\left\{x_k^{(m)}\right\}_{k\in\mathbb{N}}$, such that $\sum_{k=1}^{\infty}\left|x_k^{(m)}\right|^p<\infty$ for all $m=1,2,\ldots$. Then for each $\varepsilon>0$, there exists a positive integer N such that

$$d_p\left(x^{(m)},x^{(n)}\right)=\left(\sum_{k=1}^{\infty}\left|x_k^{(m)}-x_k^{(n)}\right|^p\right)^{\frac{1}{p}}<\varepsilon \quad \text{for all } m,n>N \tag{2.1}$$

and thus

$$\left|x_k^{(m)}-x_k^{(n)}\right|<\varepsilon \quad \text{for all } m,n>N \text{ and for all } k=1,2,\ldots.$$

This implies that for each fixed k $(1 \le k < \infty)$, $\left\{x_k^{(m)}\right\}_{m\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{K}$ $(= \mathbb{R}$ or $\mathbb{C})$. Since $\mathbb{K}$ is complete, it converges in $\mathbb{K}$. Let $x_k^{(m)} \to x_k$ as $m \to \infty$ for all k. Using these limits, we define $x = (x_1, x_2, \ldots)$ and show that $x \in \ell^p$ and $x^{(m)} \to x$.

From (2.1), we have

$$\sum_{k=1}^{l} \left|x_k^{(m)} - x_k^{(n)}\right|^p < \varepsilon^p \quad \text{for all } m,n > N \text{ and for any } l = 1,2,\ldots.$$

Letting $n \to \infty$, we obtain

$$\sum_{k=1}^{l} \left|x_k^{(m)} - x_k\right|^p \le \varepsilon^p \quad \text{for all } m > N \text{ and for any } l = 1,2,\ldots.$$

The sequence $\left\{\sum_{k=1}^{l} \left|x_k^{(m)} - x_k\right|^p\right\}_{l\ge 1}$ is a monotonically increasing sequence which is bounded above and therefore has a finite limit $\sum_{k=1}^{\infty} \left|x_k^{(m)} - x_k\right|^p$, which is less than or equal to ε^p. Therefore,

$$\sum_{k=1}^{\infty} \left|x_k^{(m)} - x_k\right|^p \le \varepsilon^p \quad \text{for all } m > N. \tag{2.2}$$

By Minkowski inequality (Theorem A.4), we have

$$\left(\sum_{k=1}^{\infty} |x_k|^p\right)^{1/p} \le \left(\sum_{k=1}^{\infty} \left|x_k^{(m)} - x_k\right|^p\right)^{1/p} + \left(\sum_{k=1}^{\infty} \left|x_k^{(m)}\right|^p\right)^{1/p}.$$

Thus $x \in \ell^p$. Moreover, from inequality (2.2), we obtain

$$d_p\left(x^{(m)}, x\right) < \varepsilon \quad \text{for all } m > N$$

which implies that $x^{(m)} \to x$ in ℓ^p. Hence ℓ^p $(1 \le p < \infty)$ is complete.

Example 2.8. The space of all bounded sequences $\ell^\infty = \left\{\{x_k\} \subseteq \mathbb{K} = \mathbb{R} \text{ or } \mathbb{C} : \sup_{1\le k<\infty} |x_k| < \infty\right\}$ with the metric $d(x,y) = \sup_{1\le k<\infty} |x_k - y_k|$, where $x = \{x_k\}, y = \{y_k\} \in \ell^\infty$, is a complete metric space.

Verification. Let $\left\{x^{(m)}\right\}_{m\in\mathbb{N}}$ be a Cauchy sequence in ℓ^∞, where $x^{(m)} = \left\{x_1^{(m)}, x_2^{(m)}, \ldots\right\} = \left\{x_k^{(m)}\right\}_{k\in\mathbb{N}}$, such that $\sup_{1\le k<\infty} \left|x_k^{(m)}\right|^p < \infty$ for all $m = 1,2,\ldots$. Then for each $\varepsilon > 0$, there exists a positive integer N such that

$$d\left(x^{(m)}, x^{(n)}\right) = \sup_{1\le k<\infty} \left|x_k^{(m)} - x_k^{(n)}\right|^p < \varepsilon \quad \text{for all } m,n > N$$

and thus

$$\left|x_k^{(m)} - x_k^{(n)}\right| < \varepsilon \quad \text{for all } m,n > N \text{ and for all } k = 1,2,\ldots. \tag{2.3}$$

This implies that for each fixed k $(1 \le k < \infty)$, $\left\{x_k^{(m)}\right\}_{m\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{K}$. Since $\mathbb{K}$ is complete, it converges in $\mathbb{K}$. Let $x_k^{(m)} \to x_k$ as $m \to \infty$ for all k. Using these limits, we define $x = (x_1, x_2, \ldots)$ and show that $x \in \ell^\infty$ and $x^{(m)} \to x$.

Letting $n \to \infty$ in (2.3), we obtain

$$\left|x_k^{(m)} - x_k\right| \le \varepsilon \quad \text{for all } m > N \text{ and for any } k = 1,2,\ldots. \tag{2.4}$$

Since $x^{(m)} = \left\{x_k^{(m)}\right\}_{k\in\mathbb{N}} \in \ell^\infty$, there is a real number N_m such that $\left|x_k^{(m)}\right| \le N_m$ for all k. Therefore,

$$\begin{aligned} |x_k| &= \left|x_k - x_k^{(m)} + x_k^{(m)}\right| \le \left|x_k^{(m)} - x_k\right| + |x_k^m| \\ &\le \varepsilon + N_m \quad \text{for all } m > N \text{ and for all } k = 1,2,\ldots. \end{aligned}$$

This inequality is true for each k and the right hand side is independent of k, it follows that $\{x_k\}$ is a bounded sequence of numbers and hence $x = \{x_k\} \in \ell^\infty$.

From (2.4), we obtain

$$d\left(x^{(m)}, x\right) = \sup_{1\le k<\infty} \left|x_k^{(m)} - x_k\right| \le \varepsilon \quad \text{for all } m > N,$$

and therefore, $x^{(m)} \to x$ in ℓ^∞. Hence ℓ^∞ is complete.

Example 2.9. The space $C[a,b]$ of all continuous real-valued functions defined on $[a,b]$ with the metric

$$d_\infty(f,g) = \max_{t\in[a,b]} |f(t) - g(t)|$$

is a complete metric space.

Verification. Let $\{f_m\}_{m\in\mathbb{N}}$ be a Cauchy sequence in $C[a,b]$. Then for each $\varepsilon > 0$, there exists a positive integer N such that

$$d_\infty(f_m, f_n) = \max_{t\in[a,b]} |f_m(t) - f_n(t)| < \varepsilon \quad \text{for all } m,n > N. \tag{2.5}$$

Then for any fixed $t_0 \in [a,b]$, we have

$$|f_m(t_0) - f_n(t_0)| < \varepsilon \quad \text{for all } m,n > N.$$

Therefore, $\{f_m(t_0)\}$ is a Cauchy sequence in $\mathbb{R}$. Since $\mathbb{R}$ is complete, this sequence converges. Let $f_m(t_0) \to f(t_0)$ as $m \to \infty$. In this way, we can associate to each $t \in [a,b]$ a unique real number $f(t)$. Thus we have defined a function f with domain $[a,b]$. Now, we show that $f \in C[a,b]$ and $f_m \to f$.

From (2.5), we obtain

$$|f_m(t) - f_n(t)| < \varepsilon \quad \text{for all } m,n > N \text{ and for all } t \in [a,b].$$

Letting $n \to \infty$, we obtain

$$|f_m(t) - f(t)| \leq \varepsilon \quad \text{for all } m > N \text{ and for all } t \in [a,b]. \tag{2.6}$$

To show that f is continuous, we consider any $t_0 \in [a,b]$ and any $\eta > 0$. According to the preceding paragraph, there exists a positive integer $N_1(\eta)$ such that

$$|f_m(t) - f(t)| < \frac{\eta}{3} \quad \text{for all } m > N_1(\eta) \text{ and for all } t \in [a,b].$$

Let $n \geq N_1(\eta)$. Then

$$|f_n(t) - f(t)| < \frac{\eta}{3} \quad \text{for all } t \in [a,b]. \tag{2.7}$$

By continuity of f_n, we obtain $\delta > 0$ such that

$$|f_n(t) - f_n(t_0)| < \frac{\eta}{3} \quad \text{whenever } |t - t_0| < \delta \tag{2.8}$$

Then from (2.7) and (2.8), we have

$$\begin{aligned} |f(t) - f(t_0)| &\leq |f(t) - f_n(t)| + |f_n(t) - f_n(t_0)| + |f_n(t_0) - f(t_0)| \\ &< \frac{\eta}{3} + \frac{\eta}{3} + \frac{\eta}{3} = \eta \end{aligned}$$

whenever $|t - t_0| < \delta$. Therefore, $f \in C[a,b]$.

Moreover, from (2.6), we have

$$\max_{t \in [a,b]} |f_m(t) - f(t)| \leq \varepsilon \quad \text{for all } m > N.$$

Thus $d_\infty(f_m, f) \leq \varepsilon$ for all $m > N$, and therefore, $f_m \to f$ as $m \to \infty$. Hence $C[a,b]$ is a complete metric space.

Example 2.10. The space $C[0,1]$ is not complete with respect to the metric

$$d_1(f,g) = \int_0^1 |f(t) - g(t)|dt \quad \text{for all } f,g \in C[0,1].$$

Verification. Let $\{f_n\}_{n \geq 2}$ be a sequence in $C[0,1]$, defined by

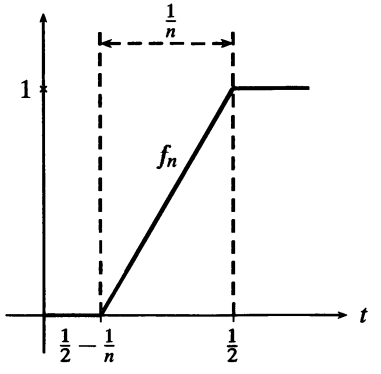
$$f_n(t) = \begin{cases} 0 & \text{if } 0 \leq t \leq \frac{1}{2} - \frac{1}{n} \\ nt - \frac{1}{2}n + 1 & \text{if } \frac{1}{2} - \frac{1}{n} < t \leq \frac{1}{2} \\ 1 & \text{if } \frac{1}{2} < t \leq 1. \end{cases}$$

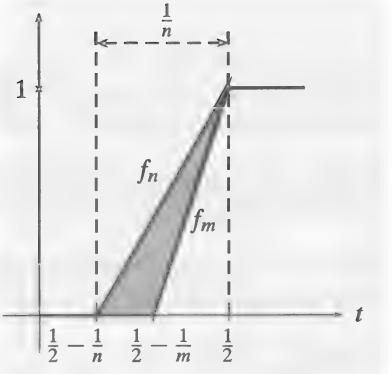
We claim that $\{f_n\}_{n \geq 2}$ is a Cauchy sequence but does not converge in $(C[0,1], d_1)$.

Indeed,

$$\begin{aligned} d(f_m, f_n) &= \int_0^1 |f_n(t) - f_m(t)|\,dt \\ &\le \int_{\frac{1}{2}-\frac{1}{m}}^{\frac{1}{2}} f_m(t)dt + \int_{\frac{1}{2}-\frac{1}{n}}^{\frac{1}{2}} f_n(t)dt \\ &= \frac{1}{2}\left(\frac{1}{m} + \frac{1}{n}\right). \end{aligned}$$

Thus, $d_1(f_n, f_m) < \frac{1}{n} + \frac{1}{m} < \varepsilon$ for all $n, m > N$, where N is a positive integer such that $N > \frac{2}{\varepsilon}$. This shows that $\{f_n\}$ is a Cauchy sequence. The measure $d_1(f_n, f_m)$ is the area of the triangle in the following figure.





$d_1(f_n, f_m)$ represents the area of the shaded triangle

Suppose that there is a function $f \in C[0,1]$ such that $d_1(f_n, f) \to 0$. But

$$d_1(f_n, f) = \int_0^{\frac{1}{2}-\frac{1}{n}} |f(t)|\,dt + \int_{\frac{1}{2}-\frac{1}{n}}^{\frac{1}{2}} |f_n(t) - f(t)|\,dt + \int_{\frac{1}{2}}^1 |1 - f(t)|\,dt. \tag{2.9}$$

Since the integrands are non-negative, so is the each integral on the right hand side of (2.9). Since $d_1(f_n, f) \to 0$ as $n \to \infty$, we have

$$\lim_{n\to\infty} \int_0^{\frac{1}{2}-\frac{1}{n}} |f(t)|dt = 0 \text{ and } \int_{\frac{1}{2}}^1 |1 - f(t)|dt = 0.$$

Since f is continuous, we have

$$f(t) = \begin{cases} 0 & \text{if } 0 \le t < \frac{1}{2} \\ 1 & \text{if } \frac{1}{2} \le t \le 1. \end{cases}$$

Therefore, f is not continuous which contradicts to our supposition that f is a continuous function. Hence $C[0,1]$ is not complete.

Example 2.11. The space of all rational numbers $\mathbb{Q}$ with the usual metric of absolute value is not complete.

Verification. Let $X = \mathbb{Q}$ be the set of all rational numbers and

$$d(x,y) = |x-y| \quad \text{for all } x,y \in X$$

be the usual metric on X. Choose a sequence $\{x_n\}$, represented in the decimal system, such that $x_n = 1.a_1a_2\cdots a_n$ is the largest rational number satisfying $x_n^2 < 2$. Then we have the sequence of numbers

$$x_1 = 1.4,\ x_2 = 1.41,\ x_3 = 1.414,\ x_4 = 1.4142,\ x_5 = 1.41421,\ldots.$$

We find that

$$d(x_m,x_n) = |x_n - x_m| = 0.00\cdots 0a_{m+1}\cdots a_n < \frac{1}{10^m} \quad \text{for all } n \geq m.$$

Therefore, we conclude that $d(x_m,x_n) \to 0$ as $m \to \infty$. So, $\{x_n\}$ is a Cauchy sequence. But this sequence converges to the irrational number $\sqrt{2} \notin \mathbb{Q}$. Thus $\{x_n\}$ does not have a limit point in $\mathbb{Q}$ and so $\mathbb{Q}$ is not complete.

Example 2.12. Let $X = (0,1]$ and d be the usual metric on X. Then (X,d) is not complete as the sequence $\left\{x_n = \frac{1}{n} : n \in \mathbb{N}\right\} \subseteq X$ is Cauchy but $x_n \to 0 \notin X$.

Exercise 2.12. Prove or disprove that $\mathbb{R}^n$ is a complete metric space with respect to the following metrics: For all $x = (x_1,x_2,\ldots,x_n)$ and $y = (y_1,y_2,\ldots,y_n) \in \mathbb{R}^n$,

(a) $d_1(x,y) = \sum_{k=1}^{n} |x_k - y_k|$;

(b) $d_p(x,y) = \left(\sum_{k=1}^{n} |x_k - y_k|^p\right)^{1/p}$ for all $p \geq 1$;

(c) $d_\infty(x,y) = \max_{1\leq k\leq n} |x_k - y_k|$.

Proof. (c) For any real numbers $x_1,x_2,\ldots,x_n$, we have

$$\left(\sum_{k=1}^{n} x_k^2\right)^{1/2} \leq \sum_{k=1}^{n} |x_k| \leq n \max_{1\leq k\leq n} |x_k| \leq n\left(\sum_{k=1}^{n} x_k^2\right)^{1/2}. \tag{2.10}$$

Indeed, the first inequality follows on squaring both the sides. The second one is obvious. The third inequality is a consequence of the fact that

$$\left(\sum_{k=1}^{n} x_k^2\right)^{1/2} \geq |x_k| \quad k = 1,2,\ldots,n.$$

Let $x = (x_1,x_2,\ldots,x_n) \in \mathbb{R}^n$ and $y = (y_1,y_2,\ldots,y_n) \in \mathbb{R}^n$. From inequality (2.10), we have

$$\left(\sum_{k=1}^{n} (x_k - y_k)^2\right)^{1/2} \leq \sum_{k=1}^{n} |x_k - y_k| \leq n \max_{1\leq k\leq n} |x_k - y_k| \leq n\left(\sum_{k=1}^{n} (x_k - y_k)^2\right)^{1/2}. \tag{2.11}$$

From inequality (2.11), we observe that a sequence in $\mathbb{R}^n$ is Cauchy (respectively, convergent) with respect to one of these metrics if and only if it is Cauchy (respectively, convergent) with respect to the other metric. Since $\mathbb{R}^n$ is complete with respect to the metric

$$d(x,y) = \left(\sum_{k=1}^{n} (x_k - y_k)^2 \right)^{1/2},$$

it is also complete with respect to the metric d_∞.

Exercise 2.13. Prove that $[0,1)$ as a subspace of the discrete metric space $\mathbb{R}$ is complete.

Exercise 2.14. Prove that the space of all natural numbers $\mathbb{N}$ with the metric

$$d(x,y) = \left| \frac{1}{x} - \frac{1}{y} \right| \quad \text{for all } x,y \in \mathbb{N},$$

is not complete

Proof. Let $\{n\}_{n\geq 1}$ be a sequence in $\mathbb{N}$. Let $\varepsilon > 0$ and N be the least integer greater than $\frac{1}{\varepsilon}$. If $m,n > N$, then

$$d(m,n) = \left| \frac{1}{m} - \frac{1}{n} \right| \leq \max\left\{ \frac{1}{m}, \frac{1}{n} \right\} < \frac{1}{N} < \varepsilon.$$

Thus, the sequence $\{n\}_{n\geq 1}$ is Cauchy. Suppose contrary that the sequence $\{n\}_{n\geq 1}$ converges to some point $p \in \mathbb{N}$. Let N_1 be any integer greater than $2p$. Then $n \geq N_1$ implies that

$$d(p,n) = \left| \frac{1}{p} - \frac{1}{n} \right| = \frac{1}{p} - \frac{1}{n} \geq \frac{1}{p} - \frac{1}{N_1} > \frac{1}{p} - \frac{1}{2p} = \frac{1}{2p}.$$

This shows that the sequence $\{n\}_{n\geq 1}$ cannot converge to p, a contradiction of our supposition. Hence $(\mathbb{N},d)$ is not complete.

Exercise 2.15. Let $X = \mathbb{R}$ and $d : X \times X \to \mathbb{R}$ be defined by

$$d(x,y) = \frac{|x-y|}{\sqrt{1+x^2}\sqrt{1+y^2}} \quad \text{for all } x,y \in X.$$

Show that (X,d) is a metric space but not complete.

Proof. We only give the proof of incompleteness. Consider the sequence $\{n\}_{n\geq 1}$ of natural numbers. Observe that

$$d(n,m) = \frac{|n-m|}{\sqrt{1+n^2}\sqrt{1+m^2}} = \frac{\left|\frac{1}{n} - \frac{1}{m}\right|}{\sqrt{\frac{1}{n^2}+1}\sqrt{\frac{1}{m^2}+1}}$$
$$\leq \left| \frac{1}{n} - \frac{1}{m} \right| \leq \frac{1}{n} + \frac{1}{m} \to 0 \text{ as } n,m \to \infty.$$

Thus $\{n\}_{n\geq 1}$ is a Cauchy sequence in (X,d). Suppose contrary that the sequence $\{n\}_{n\geq 1}$ converges to some point $p \in \mathbb{R}$. Then

$$d(n,p) = \frac{|n-p|}{\sqrt{1+n^2}\sqrt{1+p^2}} = \frac{\left|1-\frac{p}{n}\right|}{\sqrt{1+\frac{1}{n^2}}\sqrt{1+p^2}}$$
$$\to \frac{1}{\sqrt{1+p^2}} \neq 0 \text{ as } n \to \infty.$$

This shows that the sequence $\{n\}_{n\geq 1}$ does not converge to $p \in \mathbb{R}$, a contradiction. Hence (X,d) is not complete.

Exercise 2.16. Prove that the metric space $P[a,b]$ of all polynomials defined on $[a,b]$ with the uniform metric

$$d_\infty(f,g) = \max_{t\in[a,b]} |f(t)-g(t)| \quad \text{for all } f,g \in P[a,b]$$

is not complete.

Proof. Let us take $a = 0$ and $b = 1$. Consider the following sequence

$$f_n(t) = \sum_{k=0}^{n} \left(\frac{t}{2}\right)^k = 1 + \frac{t}{2} + \cdots + \frac{t^n}{2^n} \quad \text{for all } t \in [0,1].$$

Clearly, $f_n(t) \in P[0,1]$ for each $n \in \mathbb{N}$. We show that the sequence $\{f_n\}$ is Cauchy. Taking $m < n$ and observe that

$$\begin{aligned} d(f_n, f_m) &= \max_{0\leq t\leq 1} |f_n(t) - f_m(t)| \\ &= \max_{0\leq t\leq 1} \left| \sum_{k=0}^{n} \left(\frac{t}{2}\right)^k - \sum_{k=0}^{m} \left(\frac{t}{2}\right)^k \right| \\ &= \max_{0\leq t\leq 1} \left| \sum_{k=m+1}^{n} \left(\frac{t}{2}\right)^k \right| \\ &\leq \max_{0\leq t\leq 1} \left| \sum_{k=m+1}^{n} \frac{1}{2^k} \right| \\ &= \frac{1}{2^m} - \frac{1}{2^n}. \end{aligned}$$

This difference is arbitrarily small for large enough m and n, which implies that $\{f_n\}$ is a Cauchy sequence in $P[0,1]$. However, this sequence does not converge in $(P[0,1],d)$, because $\lim_{n\to\infty} f_n(t) = \dfrac{2}{2-t}$ for all $0 \leq t \leq 1$, is not a polynomial. Therefore, $(P[0,1],d)$ is not complete.

Theorem 2.7. *Let (Y,d_Y) be a subspace of a metric space (X,d). If Y is complete, then it is closed.*

Proof. Suppose that Y is a complete subspace. To prove Y is closed, it is sufficient to show that Y contains all its limit points.

Let x be a limit point of Y. Then by Theorem 2.3, there exists a sequence $\{x_n\}$ of distinct points of Y which converges to x. Since each convergent sequence is Cauchy, it is a Cauchy sequence. Also since Y is complete, the limit point x of this sequence must lie in Y. Thus Y is closed.

Theorem 2.8. *Let (X,d) be a complete metric space and (Y,d_Y) be a subspace of (X,d). Then Y is complete if and only if it is closed.*

Proof. If Y is a complete subspace of (X,d), then by Theorem 2.7, it is closed.

Conversely, assume that Y is a closed subspace of a complete metric space X. Let $\{x_n\}$ be a Cauchy sequence of points of Y. Since X is complete, this sequence converges to a point x belonging to X. By Theorem 2.3 (ii) and since Y is closed, $x \in Y$. Thus each Cauchy sequence of point of Y converges to a points of Y. Hence Y is complete.

Theorem 2.9 (Cantor's Intersection Theorem). *Let (X,d) be a complete metric space and let $\{F_n\}$ be a decreasing sequence (that is, $F_{n+1} \subseteq F_n$) of nonempty closed subsets of X such that $\delta(F_n) \to 0$ as $n \to \infty$. Then the intersection $\bigcap_{n=1}^{\infty} F_n$ contains exactly one point.*

Proof. Construct a sequence $\{x_n\}$ in X by selecting a point $x_n \in F_n$ for each $n \in \mathbb{N}$. Since $F_{n+1} \subseteq F_n$ for all n, we have $x_n \in F_n \subseteq F_m$ for all $n > m$. We claim that $\{x_n\}$ is a Cauchy sequence.

Let $\varepsilon > 0$ be given. Since $\delta(F_n) \to 0$, there exists a positive integer N such that $\delta(F_n) < \varepsilon$. Since $\{F_n\}$ is a decreasing sequence, we have $F_m, F_n \subseteq F_N$ for all $m,n \geq N$. Therefore, $x_n, x_m \in F_N$ for all $n,m \geq N$ and thus, we have

$$d(x_n,x_m) \leq \delta(F_n) < \varepsilon \quad \text{for all } n,m \geq N.$$

Hence $\{x_n\}$ is a Cauchy sequence. Since X is complete, there exists $x \in X$ such that $x_n \to x$. We claim that $x \in \bigcap_{n=1}^{\infty} F_n$.

Let n be fixed. Then the subsequence $\{x_n, x_{n+1}, \ldots\}$ of the sequence $\{x_n\}$ is contained is F_n and still converges to x, since every subsequence of a convergent sequence is convergent. But F_n being a closed subspace of the complete metric space (X,d), it is complete and so $x \in F_n$. This is true for each $n \in \mathbb{N}$. Hence $x \in \bigcap_{n=1}^{\infty} F_n$, that is, $\bigcap_{n=1}^{\infty} F_n \neq \emptyset$.

Finally, to establish that x is the only point in the intersection $\bigcap_{n=1}^{\infty} F_n$. Let $y \in \bigcap_{n=1}^{\infty} F_n$. Then x and y both are in F_n for each n. Therefore,

$$0 \leq d(x,y) \leq \delta(F_n) \to 0 \quad \text{as } n \to \infty.$$

Thus $d(x,y) = 0$ and hence $x = y$.

Remark 2.7. The assertion in Theorem 2.9 may not be true if either of the conditions

(i) each F_n is closed

(ii) $\delta(F_n) \to 0$ as $n \to \infty$

is dropped.

Example 2.13. (a) Consider the sequence $\{F_n\}$, where $F_n = [n, \infty)$ for all $n \in \mathbb{N}$, of nonempty closed subsets of the usual metric space $\mathbb{R}$ which, of course, is complete. Then $\delta(F_n) \not\to 0$ as $n \to \infty$ and $\bigcap_{n=1}^{\infty} F_n = \emptyset$.

(b) Consider the sequence $\{F_n\}$, where $F_n = \left(0, \frac{1}{n}\right)$ for all $n \in \mathbb{N}$, of nonempty subsets of the usual metric space $\mathbb{R}$. Then for each $n \in \mathbb{N}$, F_n is open but not closed, $F_{n+1} \subseteq F_n$, $\delta(F_n) \to 0$ as $n \to \infty$ and $\bigcap_{n=1}^{\infty} F_n = \emptyset$.

Now, we have the converse of Theorem 2.9.

Theorem 2.10. *Let (X,d) be a metric space. If every decreasing sequence $\{F_n\}$ of nonempty closed sets in X with $\delta(F_n) \to 0$ as $n \to \infty$ has exactly one point in its intersection, then (X,d) is complete.*

Proof. Let $\{x_n\}$ be a Cauchy sequence in X. Let

$$\begin{aligned} G_1 &= \{x_1, x_2, \ldots\}, \\ G_2 &= \{x_2, x_3, \ldots\}, \\ &\vdots \quad \vdots \\ G_n &= \{x_n, x_{n+1}, \ldots\}. \end{aligned}$$

Then $G_{n+1} \subseteq G_n$ and so $\overline{G_{n+1}} \subseteq \overline{G_n}$ for all $n \in \mathbb{N}$. Therefore, $\{\overline{G_n}\}$ is a decreasing sequence of nonempty closed sets. Since $\{x_n\}$ is a Cauchy sequence, for a given $\varepsilon > 0$, there exists a positive integer N such that

$$d(x_m, x_n) < \varepsilon \quad \text{for all } m, n > N.$$

Since $m, n > N$, we have $x_m, x_n \in G_N$ and since $d(x_m, x_n) < \varepsilon$, we have $\delta(G_N) < \varepsilon$. For $n > N$, we have $G_n \subseteq G_N$ and thus $\delta(G_n) \leq \delta(G_N) < \varepsilon$. Therefore, $\delta(G_n) \to 0$ as $n \to \infty$.

Since $\delta(G_n) = \delta(\overline{G_n})$, we have $\delta(\overline{G_n}) \to 0$ as $n \to \infty$. Taking $F_n = \overline{G_n}$, then $\{F_n\}$ is a decreasing sequence of nonempty closed sets with $\delta(F_n) \to 0$ as $n \to \infty$. Then by hypothesis, there exists an $x \in X$ such that $x \in \bigcap_{n=1}^{\infty} F_n$. Therefore, $d(x, x_n) \leq \delta(F_n)$ for all n and so $d(x, x_n) \leq \delta(F_n) \to 0$ as $n \to \infty$. Hence $x_n \to x$ in X. Thus, (X,d) is complete.

Exercise 2.17. Let (X,d) be a complete metric space and $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of open subsets of X. Then prove that

$$\operatorname{int}\left(\bigcap_{n \in \mathbb{N}} \operatorname{cl} G_n\right) = \operatorname{int} \operatorname{cl}\left(\bigcap_{n \in \mathbb{N}} G_n\right). \tag{2.12}$$

Proof. Since $\bigcap_{n\in\mathbb{N}} \operatorname{cl} G_n \supset \operatorname{cl}\left(\bigcap_{n\in\mathbb{N}} G_n\right)$, it is obvious that

$$\operatorname{int}\left(\bigcap_{n\in\mathbb{N}} \operatorname{cl} G_n\right) \supset \operatorname{int}\operatorname{cl}\left(\bigcap_{n\in\mathbb{N}} G_n\right). \tag{2.13}$$

We show the converse of inclusion (2.13). Let $x \in \operatorname{int}\left(\bigcap_{n\in\mathbb{N}} \operatorname{cl} G_n\right)$ and $r > 0$. Take $x_1 = x$. Then there exists $r_1 \in (0, r)$ such that $S_{r_1}[x_1] \subset \bigcap_{n\in\mathbb{N}} \operatorname{cl} G_n$ and therefore, $S_{r_1}[x_1] \subset \operatorname{cl} G_n$ for each $n \in \mathbb{N}$. It follows that $S_{r_1}(x_1) \cap G_1$ is an open nonempty set. Therefore, there exist $x_2 \in X$ and $r_2 \in \left(0, \frac{r_1}{2}\right]$ such that $S_{r_2}[x_2] \subset S_{r_1}(x_1) \cap G_1$. It follows that $S_{r_2}[x_2] \subset S_{r_1}[x_1] \subset \operatorname{cl} G_2$, and so $S_{r_2}(x_2) \cap G_2$ is an open nonempty set. Continuing in this way, we find the sequences $\{x_n\}_{n\in\mathbb{N}} \subset X$ and $\{r_n\}_{n\in\mathbb{N}} \subset (0,\infty)$ such that $r_n \to 0$ as $n \to \infty$, and $S_{r_{n+1}}[x_{n+1}] \subset S_{r_n}(x_n) \cap G_n$ for all n. In particular, $S_{r_{n+1}}[x_{n+1}] \subset S_{r_n}[x_n]$ for all n. Since X is a complete metric space, by Cantor's Intersection Theorem 2.9, $\bigcap_{n\in\mathbb{N}} S_{r_n}[x_n] = \{\bar{x}\}$ for some $\bar{x} \in X$. It follows that $\bar{x} \in G_n$ for all n. Since $\bar{x} \in S_{r_1}[x_1] \subset S_r(x)$, we have that $S_r(x) \cap \left(\bigcap_{n\in\mathbb{N}} G_n\right) \neq \emptyset$. As $r > 0$ is arbitrary, $x \in \operatorname{cl}\left(\bigcap_{n\in\mathbb{N}} G_n\right)$. Therefore, $\operatorname{int}\left(\bigcap_{n\in\mathbb{N}} \operatorname{cl} G_n\right) \subset \operatorname{cl}\left(\bigcap_{n\in\mathbb{N}} G_n\right)$.

Exercise 2.18. Let (X,d) be a complete metric space and $\{F_n\}_{n\in\mathbb{N}}$ be a sequence of closed subsets of X. Then prove that

$$\operatorname{cl}\left(\bigcup_{n\in\mathbb{N}} \operatorname{int} F_n\right) = \operatorname{cl}\operatorname{int}\left(\bigcup_{n\in\mathbb{N}} F_n\right). \tag{2.14}$$

Proof. It follows directly from Exercise 2.17 by taking $G_n = X \setminus F_n$ for all $n \in \mathbb{N}$.

2.4 Completion

If a metric space (X,d) is not complete, then it is always possible to construct a complete metric space which contains this incomplete metric space and some other points so that every Cauchy sequence in X has a limit point in the larger space. In fact, we need to adjoin new points to (X,d) and extend d to all these new points in such a way that the formerly nonconvergent Cauchy sequences find limits among these new points and the new points are limits of sequences in X.

Definition 2.7. Let (X,d) be a metric space. A complete metric space (X^*,d^*) is said to be a *completion* of the metric space (X,d) if X is a subspace of X^* and every point of X^* is the limit of some sequence in X.

Example 2.14. (a) The completion of the space $\mathbb{Q}$ of all rational numbers with the usual metric is the space $\mathbb{R}$ of all real numbers with the usual metric.

(b) Let $X = (0,1)$, $(0,1]$ or $[0,1)$. The completion of the metric space X with the usual metric is the metric space $[0,1]$ with the usual metric.

Definition 2.8. Let (X,d) and (X^*,d^*) be two metric spaces.

- A mapping $f : X \to X^*$ is said to be an *isometry* if

$$d^*(f(x),f(y)) = d(x,y) \quad \text{for all } x,y \in X.$$

  The mapping f is also called an *isometric embedding* of X into X^*.
- The spaces X and X^* are said to be *isometric* if there exists a bijective isometry from X to X^*.

Remark 2.8. We note that every isometry is always one-to-one.

Theorem 2.11. *Every metric space has a completion.*

Theorem 2.12. *All completions of a metric space are isometric.*

Proof. Let (X^*,d^*) and (X^{**},d^{**}) be any two completions of the metric space (X,d). Let $x^* \in X^*$ be arbitrary. By the definition of completion, there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ in X such that $x_n \to x^*$. The sequence $\{x_n\}_{n\in\mathbb{N}}$ may be assumed to belong to X^{**}. Since X^{**} is complete, $\{x_n\}_{n\in\mathbb{N}}$ converges in X^{**} to x^{**}, say. Define $f : X^* \to X^{**}$ by $f(x^*) = x^{**}$. It is clear that the mapping f is one-to-one and does not depend on the choice of the sequence $\{x_n\}_{n\in\mathbb{N}}$ converges to x^*. Moreover, by construction, $f(x) = x$ for all $x \in X$. If $x_n \to x^* \in X^*$ and $x_n \to x^{**} \in X^{**}$ while $y_n \to y^* \in X^*$ and $y_n \to y^{**} \in X^{**}$, then

$$d^*(x^*,y^*) = \lim_{n\to\infty} d(x_n,y_n) \quad \text{and} \quad d^{**}(x^{**},y^{**}) = \lim_{n\to\infty} d(x_n,y_n).$$

Thus

$$d^{**}(f(x^*),f(y^*)) = d(x^*,y^*) \quad \text{for all } x^*,y^* \in X^*.$$

Hence f is an isometry. Thus (X^*,d^*) and (X^{**},d^{**}) are isometric.

Chapter 3
Separable Spaces

It is well known that the rational numbers or irrational numbers are densely packed along the real line. This phenomenon can be characterized in terms of distance by saying that every point of $\mathbb{R}$ is zero distance from the set of rational numbers $\mathbb{Q}$ or from the set of irrational numbers $\mathbb{R} \setminus \mathbb{Q}$. The concept of closure enabled us to write this density property by the formula $\overline{\mathbb{Q}} = \mathbb{R}$ or $\overline{\mathbb{R} \setminus \mathbb{Q}} = \mathbb{R}$. In this chapter, we extend the idea of density to an arbitrary metric space. We given some other properties of a metric space, known as topological properties, such as countability, separability and Baire's category theorem.

3.1 Countability

Definition 3.1. Let $\mathscr{O}$ be the family of all open subsets of a metric space X. A subfamily $\mathscr{B} \subseteq \mathscr{O}$ of open subsets of X is said to be a *base* or *basis* for $\mathscr{O}$ if every open set $G \in \mathscr{O}$ is the union of members of $\mathscr{B}$.

Further, if the subfamily $\mathscr{B}$ is countable, then it is called *countable base* or *countable basis* for $\mathscr{O}$.

Before giving the examples of a base for a family of open sets, we mention the following characterization of a base.

Theorem 3.1. *Let $\mathscr{O}$ be the family of all open subsets of a metric space X. A subfamily $\mathscr{B} \subseteq \mathscr{O}$ is a base for $\mathscr{O}$ if and only if for every point x belonging to an open set G, there exists $B_x \in \mathscr{B}$ such that $x \in B_x \subseteq G$.*

Proof. Let $\mathscr{B}$ be a base for $\mathscr{O}$ and G be any open set in X, that is, $G \in \mathscr{O}$. Then by definition, G is the union of members of $\mathscr{B}$. Let $x \in G$. Since G in the union of members of $\mathscr{B}$, there exists a set B_x in $\mathscr{B}$ such that $x \in B_x \subseteq G$.

Conversely, let G be any arbitrary open set. Then by hypothesis, for any point $x \in G$, there exists B_x in $\mathscr{B}$ such that $x \in B_x \subseteq G$. Clearly, $G = \bigcup\{B_x \in \mathscr{B} : x \in G\}$. Then every open set is the union of members of $\mathscr{B}$.

Example 3.1. (a) The collection $\mathscr{B} = \{(a,b) : a,b \in \mathbb{R}\}$ of all open intervals forms a base for the family of all open sets in the usual metric space $\mathbb{R}$.

If the endpoints a and b are rational, then the collection $\mathscr{B} = \{(a,b) : a,b \in \mathbb{Q}\}$ of all open intervals forms a countable base for the family of all open sets in the usual metric space $\mathbb{R}$.

(b) Let X be a metric space. The collection $\{S_{r_x}(x) : x \in X \text{ and } r_x > 0\}$ of all open spheres forms a base for the family of all open sets in X.

Indeed, let G be a nonempty open subset of X and let $x \in G$. By the definition of an open set, there exists $r_x > 0$ (depending on x) such that $x \in S_{r_x}(x) \subseteq G$.

(c) Let $X = \mathbb{R}^n$ be a metric space with the metric

$$d_p(x,y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p} \quad \text{for all } p \geq 1,$$

where $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ in $\mathbb{R}^n$. Let $r > 0$ be a rational number. The collection $\{S_r(x) : x = (x_1, x_2, \ldots, x_n)$ where x_i is rational for all $i = 1, 2, \ldots, n\}$ of all spheres with rational centers and rational radii is a countable base for the family of all open sets of X.

(d) Let X be a discrete metric space. Then the collection $\mathscr{B} = \{\{x\} : x \in X\}$ forms a base for the family of all open subsets of X since every subset of a discrete metric space is open.

If X is countable, then the family $\mathscr{B}$ is also countable. In this case, $\mathscr{B}$ is a countable base for the family of all open subsets of X.

Definition 3.2. A metric space X is said to be a *second countable space* (or *second axiom space*) if there exists a countable base for the family of all open subsets of X.

Example 3.2. (a) In view of Example 3.1 (a), the collection of all open intervals (a,b) with a and b as rational points forms a countable base for the family of all open subsets of $\mathbb{R}$. Hence the usual metric space $\mathbb{R}$ is a second countable space.

(b) Let $X = \mathbb{R}^n$ be a metric space with the metric

$$d_p(x,y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p} \quad \text{for all } p \geq 1,$$

where $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ in $\mathbb{R}^n$. In view of Example 3.1 (c), (X, d_p) is a second countable space.

(c) In view of Example 3.1 (d), the discrete metric space X is second countable if X is countable.

Definition 3.3. A metric space X is said to be a *first countable space* (or *first axiom space*) if for every point $x \in X$, there exists a countable family $\{B_n(x)\}$ of open sets containing x such that every open set G containing x also contains a member of $\{B_n(x)\}$, that is, $B_n(x) \subseteq G$ for some n.

Example 3.3. The usual metric space $\mathbb{R}$ is a first countable space. Indeed, we may take $B_n(x) = \left(x - \frac{1}{n}, x + \frac{1}{n}\right)$ for each $x \in \mathbb{R}$ and $n \in \mathbb{N}$.

Remark 3.1. Every second countable metric space is first countable but converse is not true.

Example 3.4. Let X be a discrete metric space, where X is an uncountable set. Then X is first countable but not second countable.

Indeed, any open sphere $S_{1/2}(x)$ is a singleton set $\{x\}$ because X is a discrete metric space. Therefore, $\{S_{1/n}(x)\}_{n=1}^{\infty}$ is the countable family of open sets containing x such that every open set G containing x also contains a member of $\{S_{1/n}(x)\}_{n=1}^{\infty}$ because $S_{1/n}(x) = \{x\}$ for all $n = 1, 2, \ldots$. Hence X is a first countable space.

If X is uncountable, then the singleton sets $\{x\}$ are also uncountable. Hence every base is uncountable. Consequently, X is not second countable when it is uncountable.

Theorem 3.2. *Every metric space X is a first countable space.*

Proof. Let $x \in X$ and $n \in \mathbb{N}$. Set $B_n(x) = S_{1/n}(x)$. Then $\{B_n(x)\} = \{B_1(x), B_{1/2}(x), B_{1/3}(x), \ldots\}$ is a countable collection of open subsets of X such that every member of this collection contains x. Let G be an open set containing x. Then there exists $S_\varepsilon(x)$ such that $S_\varepsilon(x) \subseteq G$ for some $\varepsilon > 0$. Therefore, $B_n(x) = S_{1/n}(x) \subseteq S_\varepsilon(x) \subseteq G$ for each $n > \frac{1}{\varepsilon}$ and hence X is a first countable space.

Exercise 3.1. Let (X,d) be a metric space and Y be a nonempty subset of X. If (X,d) is second countable, then prove that (Y,d_Y) is also second countable.

Proof. Let $\mathscr{B} = \{B_n : n \in \mathbb{N}\}$ be a countable base for the family $\mathscr{O}$ of all open subsets of X. Then $\mathscr{B}^* = \{B_n \cap Y : n \in \mathbb{N}\}$ is countable. We claim that $\mathscr{B}^*$ is a base for the family $\mathscr{G}$ of all open subsets of Y.

Let G be an open set in Y and let $x \in G$. Then there exists an open set O in X such that $G = O \cap Y$. Also, $x \in O \cap Y$ and so $x \in O$ and $x \in Y$. Since $O \in \mathscr{O}$, there exists a $B_{n_0} \in \mathscr{B}$ such that $x \in B_{n_0} \subseteq O$. Since $x \in Y$, we have $x \in B_{n_0} \cap Y \subseteq O \cap Y = G$. But $B_{n_0} \cap Y \in \mathscr{B}^*$. Therefore, $\mathscr{B}^*$ forms a base for $\mathscr{G}$. Hence (Y,d_Y) is second countable.

Exercise 3.2. Let (X,d) be a metric space and Y be a nonempty subset of X. If (X,d) is first countable, is so (Y,d_Y)? Justify your answer.

Exercise 3.3. Let $X = \mathbb{R}^2$ be the usual metric space, $x = (x_1,x_2) \in X$ be any arbitrary element and $G_\lambda(x) = \{y = (y_1,y_2) \in \mathbb{R}^2 : (y_1 - x_1)^2 + 3(y_2 - x_2)^2 < \lambda\}$, where $\lambda > 0$. Prove that the family $\{G_\lambda(x) : \lambda > 0 \text{ and } x \in X\}$ is a base for the family of all open subsets of X.

Proof. Let G be any open set in X and let $x \in G$. Then there exists $r > 0$ such that $S_r(x) \subseteq G$. We know that $S_r(x) = \{y = (y_1,y_2) \in X : (y_1 - x_1)^2 + (y_2 - x_2)^2 < r\}$. Let $\lambda = r^2$. If $y \in G_\lambda(x)$, then $(y_1 - x_1)^2 + 3(y_2 - x_2)^2 < \lambda$ which implies that $(y_1 - x_1)^2 + (y_2 - x_2)^2 < \lambda = r^2$. Therefore, $y \in S_r(x)$ and so $y \in G_\lambda(x) \subseteq S_r(x) \subseteq G$. Hence $\{G_\lambda(x) : \lambda > 0 \text{ and } x \in X\}$ is a base for the family of all open subsets of X.

3.2 Dense Sets

Definition 3.4. A nonempty subset A of a metric space X is said to be *dense* (or *everywhere dense*) in X if $\overline{A} = X$, that is, if every point of X is either a point or a limit point of A.

In other words, a set A is dense in X if for any given point $x \in X$, there exists a sequence of points of A that converges to x.

Before giving the examples of dense sets, we provide a criteria for being dense.

Theorem 3.3. *Let A be a nonempty subset of a metric space X. The following statements are equivalent.*

(a) For every $x \in X$, $\rho(x,A) = 0$.
(b) $\overline{A} = X$.
(c) A has nonempty intersection with every nonempty open subset of X.

Proof. The equivalence of (a) and (b) directly follows from Theorem 1.5. We prove the equivalence of (b) and (c).

(b) $\Rightarrow$ (c). Suppose contrary that there exists a nonempty open subset O of X such that $A \cap O = \emptyset$. Then $A \subseteq O^c$ and so $\overline{A} \subseteq \overline{O^c} = O^c$ because O^c is closed. Therefore, $\overline{A} \subseteq O^c$ whence $\overline{A} \cap O = \emptyset$ and thus $\overline{A} \neq X$, a contradiction of (b). Hence (c) holds.

(c) $\Rightarrow$ (b). Suppose that (c) holds but (b) does not, that is, $\overline{A} \neq X$. Then $(\overline{A})^c \neq \emptyset$. Since $\overline{A}$ is a closed set, $(\overline{A})^c$ is an open set in X and disjoint from A, that is, $(\overline{A})^c \cap A = \emptyset$. Since $(\overline{A})^c$ is nonempty open subset of X, by hypothesis, $(\overline{A})^c \cap A \neq \emptyset$, a contradiction. Hence $\overline{A} = X$.

Remark 3.2. It can be easily seen that a subset A of X is dense if and only if A^c has empty interior.

Example 3.5. (a) The set of all rational numbers $\mathbb{Q}$ is dense in the usual metric space $\mathbb{R}$ since $\overline{\mathbb{Q}} = \mathbb{R}$.

(b) Since $\overline{\mathbb{R} \setminus \mathbb{Q}} = \mathbb{R}$, the set of all irrational numbers $\mathbb{R} \setminus \mathbb{Q}$ is dense in the usual metric space $\mathbb{R}$.

(c) The set $A = \{a + ib \in \mathbb{C} : a, b \in \mathbb{Q}\}$ is dense in $\mathbb{C}$ since $\overline{A} = \mathbb{C}$.

(d) The set $\mathbb{Q}^n = \underbrace{\mathbb{Q} \times \mathbb{Q} \times \cdots \times \mathbb{Q}}_{n \text{ times}}$ is dense in $\mathbb{R}^n$ with the usual metric.

(e) The set

$$A = \{x = (a_1, a_2, \ldots, a_n, 0, 0, \ldots) : a_i \in \mathbb{Q} \text{ for all } 1 \leq i \leq n \text{ and } n \in \mathbb{N}\}$$

is dense in the space $\ell^p, 1 \leq p < \infty$, with the following metric

$$d_p(x,y) = \left(\sum_{i=1}^{\infty} |x_i - y_i|^p \right)^{1/p},$$

where $x = \{x_1, x_2, \ldots\}$ and $y = \{y_1, y_2, \ldots\}$ in ℓ^p.

Indeed, we have to show that $\overline{A} = \ell^p$. Let $x = \{x_1, x_2, \ldots\} \in \ell^p$ be an arbitrary element and let $\varepsilon > 0$ be given. Then there exists a positive integer N such that

$$\sum_{i=N+1}^{\infty} |x_i|^p < \frac{\varepsilon^p}{2}.$$

Since $\mathbb{Q}$ is dense in $\mathbb{R}$, for every real number, there is a rational number which is as close to it as we wish. Consequently, we can choose an element $y = (a_1, a_2, \ldots, a_N, 0, 0, \ldots) \in A$ such that

$$\sum_{i=1}^{N} |x_i - a_i|^p < \frac{\varepsilon^p}{2}.$$

Hence

$$[d(x,y)]^p = \sum_{i=1}^{\infty} |x_i - a_i|^p$$
$$= \sum_{i=1}^{N} |x_i - a_i|^p + \sum_{i=N+1}^{\infty} |x_i|^p$$
$$< \frac{\varepsilon^p}{2} + \frac{\varepsilon^p}{2} = \varepsilon^p,$$

and this implies that

$$d(x,y) < \varepsilon.$$

It means that the distance from $x \in \ell^p$ to A is zero. By Theorem 3.3, A is dense in ℓ^p.

(f) The set $P[a,b]$ of all polynomials defined on $[a,b]$ with rational coefficients is dense in $C[a,b]$.

Indeed, let $f \in C[a,b]$. By Weierstrass's Theorem (Theorem A.6), there exists a sequence $\{p_n\}$ of polynomials with real coefficients that converges uniformly to f on $[a,b]$, that is, for each $\varepsilon > 0$, there exists a positive integer N such that for all $t \in [a,b]$

$$|p_n(t) - f(t)| < \varepsilon \quad \text{whenever} \quad n > N$$

and so $d_\infty(p_n, f) < \varepsilon$ for all $n > N$. But any polynomial can be uniformly approximated by a polynomial with rational coefficients since $\mathbb{Q}$ is dense in $\mathbb{R}$. Therefore, corresponding to the sequence $\{p_n\}$, there exists a sequence $\{q_n\}$ of polynomials with rational coefficients (that is, $\{q_n\} \in P[a,b]$) such that

$$|p_n(t) - q_n(t)| < \frac{1}{n} \quad \text{for all } t \in [a,b].$$

This implies that

$$d_\infty(p_n, q_n) = \sup_{t \in [a,b]} |p_n(t) - q_n(t)| < \frac{1}{n} \quad \text{for all } n = 1,2,\ldots.$$

By the triangle inequality, we have

$$d_\infty(f, q_n) \le d_\infty(f, p_n) + d_\infty(p_n, q_n) \to 0 \text{ as } n \to \infty.$$

Hence $q_n \to f \in C[a,b]$ as $n \to \infty$ and thus $\overline{P[a,b]} = C[a,b]$.

(g) Let (X,d) be a discrete metric space. Since every subset of X is closed, the only dense subset of X is itself.

(h) By the definition of a completion, every metric space is dense in its completion.

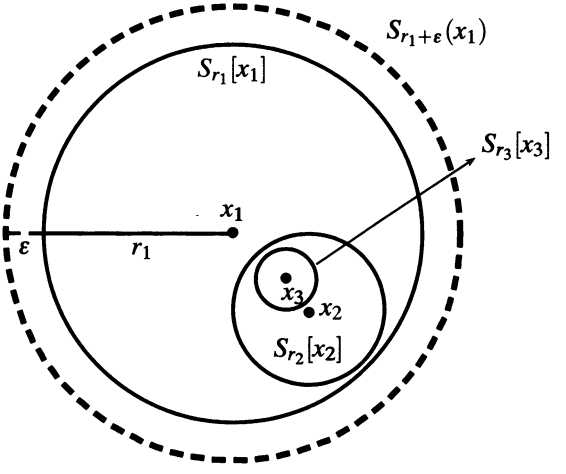
Theorem 3.4. *Let $\{G_n\}$ be a sequence of dense open sets in a complete metric space X. Then $\bigcap_{n=1}^{\infty} G_n$ is also dense in X.*

Proof. Let S be any open sphere in X. In view of Theorem 3.3, it is sufficient to show that $\left(\bigcap_{n=1}^{\infty} G_n\right) \cap S \neq \emptyset$. Since each G_n is dense in X, $\overline{G_n} = X$ and $G_n \cap S$ is

nonempty by Theorem 3.3 and open as the intersection of two open sets. In particular, $G_1 \cap S$ is a nonempty open set. Therefore, there exist $x_1 \in G_1 \cap S$ and $r_1 > 0$ such that $S_{r_1}[x_1] \subset S_{r_1+\varepsilon}(x_1) \subseteq G_1 \cap S$ for $\varepsilon > 0$. Similarly, $G_2 \cap S_{r_1}(x_1)$ is nonempty and open, so there exist $x_2 \in G_2 \cap S_{r_1}(x_1)$ and $r_2 > 0$ with $r_2 \leq \frac{r_1}{2}$ such that $S_{r_2}[x_2] \subseteq G_2 \cap S_{r_1}(x_1)$. Continue in this way, we obtain a sequence $\{x_n\}$ such that

$$S_{r_n}[x_n] \subseteq G_n \cap S_{r_{n-1}}(x_{n-1}) \quad \text{for all } n = 1, 2, \ldots,$$

with $r_n \leq \frac{r_{n-1}}{2} = \frac{r_1}{2^{n-1}} \to 0$ as $n \to \infty$.



Observe that whenever $m > n$, we have

$$S_{r_1}[x_1] \supseteq \cdots \supseteq S_{r_n}[x_n] \supseteq \cdots \supseteq S_{r_m}[x_m].$$

Note that X is a complete metric space and $\{S_{r_n}[x_n]\}$ is a decreasing sequence of nonempty closed subsets of X such that $\delta(S_{r_n}[x_n]) \to 0$ as $n \to \infty$. By Cantor's Intersection Theorem 2.9, $\bigcap_{n=1}^{\infty} S_{r_n}[x_n]$ contains exactly one point, say, x. Then

$$\{x\} = \bigcap_{n=1}^{\infty} S_{r_n}[x_n] \subseteq \bigcap_{n=1}^{\infty} G_n.$$

Since x is also in S, we have $\left(\bigcap_{n=1}^{\infty} G_n\right) \cap S \neq \emptyset$. Hence $\bigcap_{n=1}^{\infty} G_n$ is dense in X.

Definition 3.5. A metric space X is said to be *separable* if there exists a countable dense set in X.

A metric space which is not separable is called *inseparable*.

Example 3.6. (a) The usual metric space $\mathbb{R}$ is separable since the set of all rational numbers $\mathbb{Q}$ is dense in $\mathbb{R}$.

(b) The usual metric space $\mathbb{C}$ is separable since the set $A = \{a+ib \in \mathbb{C} : a,b \in \mathbb{Q}\}$ is dense in $\mathbb{C}$.

(c) The Euclidean space $\mathbb{R}^n$ is separable since the set $\mathbb{Q}^n = \underbrace{\mathbb{Q} \times \mathbb{Q} \times \cdots \times \mathbb{Q}}_{n \text{ times}}$ is countable and dense in $\mathbb{R}^n$.

(d) The space $\ell^p, 1 \leq p < \infty$, is separable as the set

$$A = \{x = (a_1, a_2, \ldots, a_n, 0, 0, \ldots) : a_i \in \mathbb{Q}, 1 \leq i \leq n \text{ and for all } n \in \mathbb{N}\}$$

is countable and dense in the space ℓ^p.

Indeed, if A_n denotes the subset of all those elements $x = \{a_i\}_{i \geq 1}$ such that $a_j = 0$ for all $j \geq n+1$, then A_n is countable and so $A = \bigcup_{n=1}^{\infty} A_n$. By Example 3.5 (e), A is dense in the space ℓ^p.

(e) The space $C[a,b]$ is separable since the set $P[a,b]$ of all polynomials defined on $[a,b]$ with rational coefficients is countable and dense in $C[a,b]$ by Example 3.5 (f).

(f) A discrete metric space X is separable if and only if the set X is countable.

Example 3.7. The space ℓ^∞ of all bounded sequences of real or complex numbers with the metric

$$d(x,y) = \sup_{1 \leq n < \infty} |x_n - y_n|$$

where $x = \{x_n\}$ and $y = \{y_n\}$ in ℓ^∞, is not separable.

Consider a subset $A = \{x = \{x_n\} \subseteq \ell^\infty : x_n = 0 \text{ or } 1\}$ of ℓ^∞. We claim that A is uncountable.

If B is any countable subset of A, then the elements of B can be arranged in a sequence $s_1, s_2, \ldots$. We construct a sequence s as follows. If the mth element of s_m is 1, then the mth element of s is 0, and vice versa. Then the element s of X differs from each s_m in the mth place and is therefore equal to none of them. So, $s \notin B$ although $s \in A$. This shows that any countable subset of A must be a proper subset of A. It follows that A must be uncountable, for if it were to be countable, then it would have to be a proper subset of itself, which is absurd. We proceed to use the uncountability of the subset A to argue that ℓ^∞ must be inseparable.

The distance between two distinct elements x and y of A is $d(x,y) = \sup_{1 \leq n < \infty} |x_n - y_n| = 1$. Suppose contrary that E is a countable dense subset of ℓ^∞. Consider the spheres of radii $\frac{1}{3}$ whose centers at the points of E. Then their union is the entire space ℓ^∞, because E is dense, and in particular contains A. Since the spheres are countable in numbers while A is not, in at least one sphere there must be two distinct elements x and y of A. Let x_0 denote the center of such a sphere. Then

$$1 = d(x,y) \leq d(x,x_0) + d(x_0,y) < \frac{1}{3} + \frac{1}{3} < 1,$$

which is impossible. Hence (ℓ^∞, d) is not separable.

Theorem 3.5. *Let (X,d) be a metric space and Y be a nonempty subset of X. If X is separable, then so is Y with respect to the induced metric.*

Proof. Let $S = \{x_n : n \in \mathbb{N}\}$ be a countable dense subset of X. If S is contained in Y, then there is nothing to prove. Otherwise, we construct a countable dense subset of Y whose points are arbitrary close to those of S. For positive integers m and k, let $S_{m,k} = S_{\frac{1}{k}}(x_m)$ and choose $y_{m,k} \in S_{m,k} \cap Y$ whenever this set is nonempty. Then $d(x_m, y_{m,k}) < \frac{1}{k}$. We claim that the countable subset $A = \{y_{m,k} : m \text{ and } k \text{ are positive integers}\}$ of Y is dense in Y.

Let $y \in Y$ and $\varepsilon > 0$ be given. Let k be so large that $\frac{1}{k} < \frac{\varepsilon}{2}$ and find $x_m \in S_{\frac{1}{k}}(y)$. Then $d(y, x_m) < \frac{1}{k}$ and $y \in S_{m,k} \cap Y$. Thus

$$d(y, y_{m,k}) \leq d(y, x_m) + d(x_m, y_{m,k}) < \frac{1}{k} + \frac{1}{k} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Therefore, $y_{m,k} \in S_\varepsilon(y)$. Since $y \in Y$ and ε were arbitrary, it follows that $y \in \overline{A}$ and hence $\overline{A} = Y$. Thus A is countable dense in Y.

Theorem 3.6. *A metric space (X, d) is separable if and only if it is second countable.*

Proof. Let (X, d) be a separable space and $A = \{a_n : n \in \mathbb{N}\}$ be a countable dense subset of X. Consider the countable collection of open spheres $\mathscr{B} = \{S_{1/k}(a_j) : j, k \in \mathbb{N}\}$ with centers at a_j, $j = 1, 2, \ldots$ and radii $r_k = \frac{1}{k}$ are rational for all $k = 1, 2, \ldots$, that is,

$$\mathscr{B} = \{S_{r_i}(a_j) : a_j \in A \text{ for } j = 1, 2, \ldots \text{ and } r_i \text{ is a rational number for } i = 1, 2, \ldots\}.$$

Then we claim that $\mathscr{B}$ is a base for the family of all open sets in X. Let G be any open set in X and $x \in G$. Then for some $r > 0$, $S_r(x) \subseteq G$. Choose a positive integer m such that $\frac{1}{m} < \frac{r}{2}$. Since A is dense in X, there exists $a_j \in A$ such that $a_j \in S_{1/m}(x)$, that is, $d(x, a_j) < \frac{1}{m}$. We now show that $S_{1/m}(a_j) \subseteq S_r(x)$. Let $y \in S_{1/m}(a_j)$, then $d(a_j, y) < \frac{1}{m}$. By the triangle inequality, we have

$$d(x, y) \leq d(x, a_j) + d(a_j, y) < \frac{1}{m} + \frac{1}{m} = \frac{2}{m} < r.$$

This implies that $y \in S_r(x)$. Hence $S_{1/m}(a_j) \subseteq S_r(x) \subseteq G$. Also, $x \in S_{1/m}(a_j)$. Thus, for every $x \in G$, there exists $S_{1/m}(a_j) \in \mathscr{B}$ containing x and contained in G. Hence $\mathscr{B}$ is a countable base for the family of all open sets in X and thus X is a second countable space.

Conversely, assume that X is a second countable space. Let $\mathscr{B} = \{B_n : n \in \mathbb{N}\}$ be a countable base for the family of all open subsets of X. For each $n \in \mathbb{N}$, we choose a point $b_n \in B_n$. Let $A = \{b_n \in B_n : n \in \mathbb{N}\}$. Then A is countable. We show that $\overline{A} = X$. Let $x \in X$ be arbitrary and G be any open set containing x. Since $\mathscr{B}$ is a base, there exists at least one $B_{n_0} \in \mathscr{B}$ such that $x \in B_{n_0} \subseteq G$. From the definition of A, we have $b_{n_0} \in A$ for all $b_{n_0} \in B_{n_0}$. Therefore, G contains a point of A other than x and thus, x is a limit point of A. Hence $x \in \overline{A}$ and so $\overline{A} = X$.

Exercise 3.4. Let A a nonempty subset of a metric space X. Prove that the following statements are equivalent.

(a) A is dense in X.
(b) The only closed superset of A is X.
(c) The only open set disjoint from A is $\emptyset$.

(d) A intersects every nonempty open set.
(e) A intersects every open sphere.

Exercise 3.5. Let $\{G_n\}$ be a sequence of dense open sets in a complete metric space X. Prove that $\bigcup_{n\in\mathbb{N}} G_n$ is dense in X.
Hint: The proof lies on the lines of the proof of Theorem 3.4.

Exercise 3.6. Let (X,d) be a metric space and Y be a subset of X. If Y is separable and $\overline{Y} = X$, then prove that X is separable.

Proof. Let A be a countable dense subset of Y. Let $x \in X$ and $\varepsilon > 0$ be given. Since $\overline{Y} = X$, there exists $y \in Y$ such that $d(x,y) < \frac{\varepsilon}{2}$. Also, since A is dense in Y, there exists $z \in A$ such that $d(y,z) < \frac{\varepsilon}{2}$. By the triangle inequality, we have

$$d(x,z) \leq d(x,y) + d(y,z) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

It follows that A is dense in X and hence X is separable.

3.3 Nowhere Dense Sets

Definition 3.6. A subset A of a metric space X is said to be *nowhere dense* in X if $(\overline{A})^\circ = \emptyset$, that is, $\overline{A}$ contains no interior point.

Example 3.8. (a) In the usual metric space $\mathbb{R}$, any finite set is nowhere dense.

(b) The sets $\mathbb{N}$ and $\mathbb{Z}$ are nowhere dense in $\mathbb{R}$ since $\overline{\mathbb{N}} = \mathbb{N}$ and $\mathbb{N}^\circ = \emptyset$.

(c) In the usual metric space $\mathbb{R}^2$, the set of points on a line is nowhere dense.

(d) In the discrete metric space, the empty set is the only nowhere dense set.

Remark 3.3. The notion of nowhere dense is not the opposite of everywhere dense, that is, if a set is not nowhere dense then it does not imply that the set is everywhere dense. For example, consider the usual metric space $\mathbb{R}$ and the set $A = \{x \in \mathbb{R} : 1 < x < 2\}$. Since

$$(\overline{A})^\circ = A \neq \emptyset \text{ and } (\overline{A})^c = \{x \in \mathbb{R} : x < 1 \text{ or } x > 2\} = (-\infty, 1) \cup (2, \infty),$$

A is not nowhere dense as well as not everywhere dense.

Theorem 3.7. *Let A be a subset of a metric space X. Then A is nowhere dense in X if and only if $X \setminus \overline{A}$ is dense in X.*

Proof. Since $\overline{X \setminus A} = X \setminus A^\circ$, we have $A^\circ = X \setminus \left(\overline{X \setminus A}\right)$. Replacing A by $\overline{A}$, we get $(\overline{A})^\circ = X \setminus \overline{(X \setminus \overline{A})}$. Therefore, $(\overline{A})^\circ = \emptyset$ if and only if $X = \overline{(X \setminus \overline{A})}$.

Corollary 3.1. *Let A be a closed subset of a metric space X. Then A is nowhere dense in X if and only if $X \setminus A$ is dense in X.*

Remark 3.4. For any subset A of a metric space X, we have $X \setminus \overline{A} \subseteq X \setminus A$. Therefore from Theorem 3.7, it follows that if A is nowhere dense in X, then $X \setminus A$ is dense in X. However, the converse is not true. For example, the set of all irrational numbers $\mathbb{R} \setminus \mathbb{Q}$ is dense in the usual metric space $\mathbb{R}$. However $\mathbb{Q}$ is not nowhere dense set in $\mathbb{R}$ because $\overline{\mathbb{Q}} = \mathbb{R}$.

Definition 3.7. A subset A of a metric space X is said to be of *first category* if it is a countable union of nowhere dense subsets. The set A is said to be of *second category* if it is not of first category.

Remark 3.5. (i) Every subset of a metric space must be either of first category or of second category.

(ii) If $\{A_n\}$ be a countable family of first category sets, then $\bigcup_n A_n$ is of first category because a countable union of countable sets is also countable.

(iii) If X is a metric space of second category and $A_1 \subseteq X$ is of first category such that $X = A_1 \cup A_2$, then A_2 is of second category.

(iv) Since a subset of a nowhere dense set is nowhere dense, and therefore, a subset of a set of first category is of first category.

Example 3.9. (a) The empty set is of first category.

(b) The set of all rational numbers $\mathbb{Q}$ in $\mathbb{R}$ is of first category.

Indeed, if $x_1, x_2, \ldots$ is an enumeration of the rational numbers, then each singleton set $\{x_i\}$ is closed and $\{x_i\}^\circ = \emptyset$. It follows that $\bigcup\{x_i\}$, the set of all rational numbers in $\mathbb{R}$, is of first category.

Theorem 3.8 (Baire's Category Theorem). *Every complete metric space is of second category.*

Proof. Let $\{A_n\}$ be a sequence of nowhere dense sets in a complete metric space X such that $X = \bigcup_{n=1}^{\infty} A_n$. Since A_n is nowhere dense, $\overline{A_n}$ has no interior points, so any open sphere in X must intersects $G_n = X \setminus \overline{A_n}$ for all $n = 1, 2, \ldots$. Therefore, $\{G_n\}$ is a family of dense open sets in X. By Theorem 3.4, $\bigcap_{n=1}^{\infty} G_n \neq \emptyset$. It follows that $\bigcup_{n=1}^{\infty} A_n \neq X$, which is a contradiction.

Remark 3.6. Since $\mathbb{R}$ is complete, by Baire's category theorem, $\mathbb{R}$ cannot be written as a countable union of nowhere dense sets.

Exercise 3.7. Let A be a subset of a metric space X. Prove that the following statements are equivalent.

(a) A is nowhere dense in X.
(b) $\overline{A}$ does not contain any nonempty open set.
(c) Every nonempty open set has a nonempty open subset disjoint from $\overline{A}$.
(d) Every nonempty open set contains a nonempty open subset disjoint from A.
(e) Every nonempty open set contains an open sphere disjoint from A.

Exercise 3.8. Prove that the Cantor set is nowhere dense.

Exercise 3.9. Let A be an open subset of a metric space X. Prove that A is dense in X if and only if $X \setminus A$ is nowhere dense in X.

Exercise 3.10. Prove that a finite union of nowhere dense sets in a metric space is a nowhere dense set.

Proof. Let A_1 and A_2 be nowhere dense subsets of a metric space X and let G be an arbitrary open set in X. Then by Exercise 3.7 (d), there exists an open set $G_1 \subseteq G$ such that $G_1 \cap A_1 = \emptyset$. Similarly, there exists an open set $G_2 \subseteq G_1 \subseteq G$ such that $G_2 \cap A_2 = \emptyset$. Since

$$G_2 \cap (A_1 \cup A_2) = (G_2 \cap A_1) \cup (G_2 \cap A_2) = \emptyset,$$

G contains a nonempty open set G_2 such that G and $A_1 \cup A_2$ are disjoint. Again, by Exercise 3.7 (d), $A_1 \cup A_2$ is nowhere dense. By induction, one can easily prove that the finite union of nowhere dense sets is nowhere dense.

Exercise 3.11. Give an example to show that a countable (infinite) union of nowhere dense sets in a metric space X need not be a nowhere dense set in X.

Solution. Let $X = \mathbb{R}$ be a metric space with the usual metric. Consider the set $\mathbb{Q}$ of all rational numbers as a subset of X. We can write

$$\mathbb{Q} = \bigcup \left\{ \left\{ \frac{p}{q} \right\} : p \in \mathbb{Z}, q \in \mathbb{N} \right\}.$$

Then $\mathbb{Q}$ has been written as a countable union of nowhere dense sets $\left\{ \frac{p}{q} \right\}$ in $\mathbb{R}$. It is well known that $\mathbb{Q}$ is not nowhere dense set in $\mathbb{R}$ as $\mathbb{Q}$ is everywhere dense set in $\mathbb{R}$.

Exercise 3.12. If $\{F_n\}$ is a sequence of closed subsets of the complete metric space X such that $X = \bigcup_{n=1}^{\infty} F_n$, then prove that at least one of F_n's has nonempty interior.

Hint: Follows from Exercise 2.18.

Exercise 3.13. Prove that the subset $\mathbb{R} \setminus \mathbb{Q}$, the set of all irrational numbers, of $\mathbb{R}$ is of second category.

Proof. Since $\mathbb{R}$ is a complete metric space, by Baire's category Theorem 3.8, $\mathbb{R}$ is of second category. By Example 3.9 (b), the set $\mathbb{Q}$ of all rational numbers is of first category. Since $\mathbb{R} = \mathbb{Q} \cup (\mathbb{R} \setminus \mathbb{Q})$, by Remark 3.5 (iii), $\mathbb{R} \setminus \mathbb{Q}$ is of second category.

This page intentionally left blank

Chapter 4
Compact Spaces

It is well known that every sequence in a closed and bounded subset of $\mathbb{R}$ has a subsequence converges in it. This result is not true in general metric spaces. The metric spaces in which every sequence has a convergent subsequence are called sequentially compact. We present the definition of a compact set in a metric space and its characterizations. The Bolzano-Weierstrass property is given in the setting of a metric space. We prove that three concepts, namely, compactness, sequentially compactness and Bolzano-Weierstrass property are equivalent in the setting of a metric space.

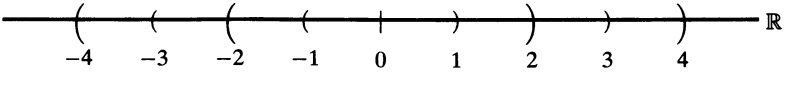
4.1 Definitions and Basic Concepts

Definition 4.1. Let X be a metric space and Λ be any index set.

- A collection $\mathscr{F} = \{G_\alpha\}_{\alpha\in\Lambda}$ of subsets of X is called a *cover* of X if $\bigcup_{\alpha\in\Lambda} G_\alpha = X$, that is, every element of X belongs to at least one member of $\mathscr{F}$. If each member of $\mathscr{F}$ is an open set in X then it is called an *open cover* of X.
- A subcollection $\mathscr{C}$ of a cover $\mathscr{F}$ of X is called a *subcover* if $\mathscr{C}$ is itself a cover of X. $\mathscr{C}$ is called a *finite subcover* if it consists only a finite number of members.
  In other words, if there exist $G_{\alpha_1}, G_{\alpha_2}, \ldots, G_{\alpha_n} \in \mathscr{F}$ such that $\bigcup_{k=1}^{n} G_{\alpha_k} = X$, then the subcollection $\mathscr{C} = \{G_{\alpha_1}, G_{\alpha_2}, \ldots, G_{\alpha_n}\}$ is called a finite subcover of X.
  In this case, $\mathscr{F}$ is said to be *reducible to a finite cover* or contains a *finite subcover*.

Definition 4.2. Let X be a metric space and Y be a subset of X. A collection $\mathscr{F} = \{G_\alpha\}_{\alpha\in\Lambda}$ of subsets of X is said to *cover* Y if $Y \subseteq \bigcup_{\alpha\in\Lambda} G_\alpha$.

Example 4.1. (a) Let $\mathbb{R}$ be the usual metric space, $\mathscr{F} = \{(-n,n) : n \in \mathbb{N}\}$ and $\mathscr{C} = \{(-2n,2n) : n \in \mathbb{N}\}$. Then $\mathscr{F}$ is an open cover of $\mathbb{R}$ and $\mathscr{C}$ is a subcover of $\mathscr{F}$.

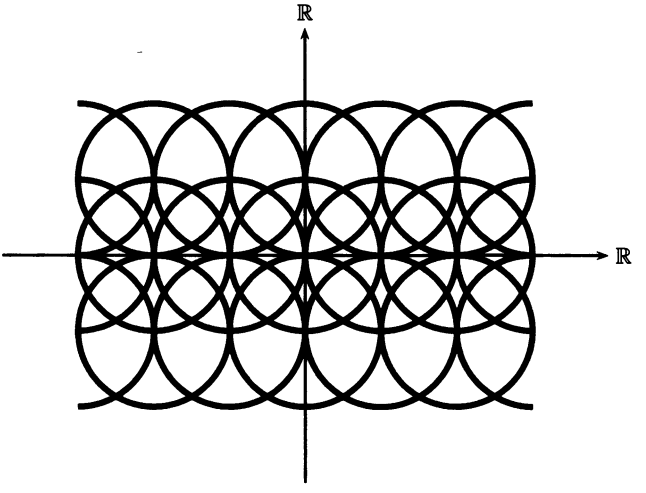


(b) Let $(0,1)$ be a metric space with the usual metric. Let

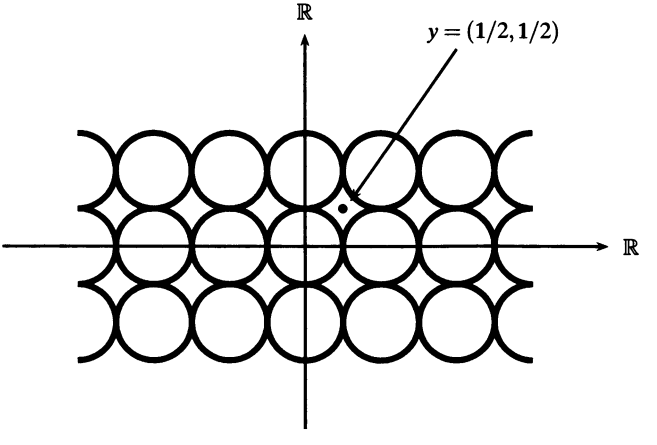
$$\mathscr{F} = \left\{\left(0, 1 - \frac{1}{n+1}\right) : n \in \mathbb{N}\right\} \text{ and } \mathscr{C} = \left\{\left(0, 1 - \frac{1}{4(n+1)}\right) : n \in \mathbb{N}\right\}.$$

Then $\mathscr{F}$ is an open cover of $(0,1)$ and $\mathscr{C}$ is a subcover of $\mathscr{F}$.

(c) Let $\mathbb{R}^2$ be the usual metric space. Consider the family $\mathscr{F} = \{S_1(x) : x \in \mathbb{Z} \times \mathbb{Z}\}$ of open spheres $S_1(x)$ in the plane $\mathbb{R}^2$ with radius 1 and center $x = (m,n)$, where m and n are integers. Then $\mathscr{F}$ is a cover of $\mathbb{R}^2$, that is, every point in $\mathbb{R}^2$ belongs to at least one member of $\mathscr{F}$.



On the other hand, the family $\mathscr{C} = \{S_{1/2}(x) : x \in \mathbb{Z} \times \mathbb{Z}\}$ of open spheres $S_{1/2}(x)$ in the plane $\mathbb{R}^2$ with radius 1/2 and center $x = (m,n)$, where m and n are integers, is not a cover of $\mathbb{R}^2$ because the point $(1/2,1/2) \in \mathbb{R}^2$ does not belong to any member of $\mathscr{C}$.



Definition 4.3. A metric space X is said to be *compact* if every open cover of X has a finite subcover.

A subspace (Y, d_Y) of a metric space (X, d) is called *compact* if it is compact as a metric space in its own right.

A subset Y of a metric space X is called *compact* if it is compact as a metric subspace. In other words, a nonempty subset Y of a metric space (X, d) is compact if it is a compact metric space with the metric induced on it by d.

Example 4.2. (a) The usual metric space $\mathbb{R}$ is not compact.

Consider an open cover $\mathscr{F} = \{(-n, n) : n \in \mathbb{N}\}$ of $\mathbb{R}$ by open intervals. Let $\{(-n_i, n_i) : 1 \le i \le k\}$ be any finite subcollection of $\mathscr{F}$ and let $n^* = \max\{n_1, n_2, \ldots, n_k\}$. Then $n^* \notin \bigcup_{i=1}^{k} (-n_i, n_i)$. Thus no finite subcollection of $\mathscr{F}$ covers $\mathbb{R}$. Hence $\mathbb{R}$ is not compact.

(b) Any finite subset Y of a metric space X is compact, because in this case every open cover of Y is finite.

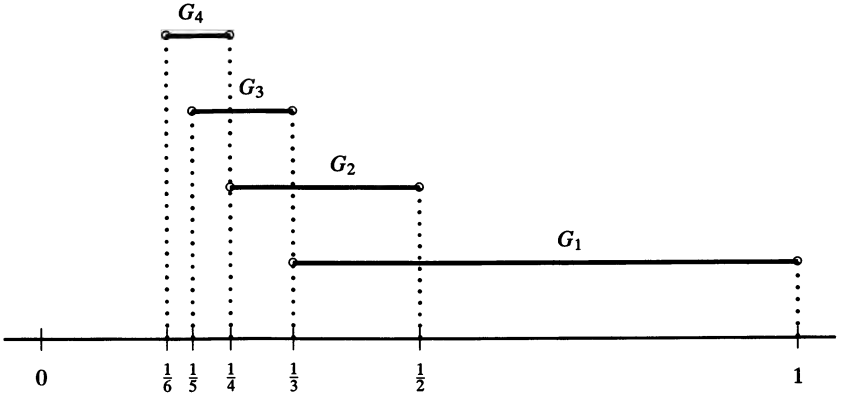
(c) The set of integers $\mathbb{Z}$ with the usual metric $d(x, y) = |x - y|$ for all $x, y \in \mathbb{Z}$ is not a compact metric space.

Indeed, $\{n\} = \mathbb{Z} \cap \left(n - \frac{1}{2}, n + \frac{1}{2}\right)$ is an open cover of $\mathbb{Z}$. But this open cover does not have any finite subcover. Thus $\mathbb{Z}$ is not compact.

(d) The subspace $X = \{0\} \cup \left\{\frac{1}{n} : n \in \mathbb{N}\right\}$ of $\mathbb{R}$ is compact.

Indeed, if $\mathscr{F}$ is a given open cover of X, then there is an element G of $\mathscr{F}$ containing 0. The set G contains all but finitely many of the points $\frac{1}{n}$. For each point of X not in G, choose a member of $\mathscr{F}$ containing it. The collection consisting of these members of $\mathscr{F}$ along with the member G, is a finite subcollection of $\mathscr{F}$ that covers X.

(e) Let $\mathbb{R}$ be the usual metric space and $Y = (0, 1)$ be the subspace of $\mathbb{R}$. Then Y is not compact.



Consider the class of open intervals

$$\mathscr{F} = \left\{\left(\frac{1}{3}, 1\right), \left(\frac{1}{4}, \frac{1}{2}\right), \left(\frac{1}{5}, \frac{1}{3}\right), \left(\frac{1}{6}, \frac{1}{4}\right), \ldots\right\}.$$

Then $Y = \bigcup_{n=1}^{\infty} G_n$, where $G_n = \left(\frac{1}{n+2}, \frac{1}{n}\right)$ for all $n \in \mathbb{N}$. Hence $\mathscr{F}$ is an open cover of Y. But $\mathscr{F}$ contains no finite subcover.

Indeed, let $\mathscr{F}^* = \{(a_1,b_1),(a_2,b_2),\ldots,(a_n,b_n)\}$ be any finite subcover of $\mathscr{F}$. If $\varepsilon = \min\{a_1,a_2,\ldots,a_n\}$, then $\varepsilon > 0$ and $(a_1,b_1) \cup (a_2,b_2) \cup \cdots \cup (a_n,b_n) \subset (\varepsilon,1)$. But $(0,\varepsilon]$ and $(\varepsilon,1)$ are disjoint. Hence $\mathscr{F}^*$ is not a cover of Y and so Y is not compact.

(f) Let $\mathbb{R}$ be the usual metric space and $Y = (0,1]$ be the subspace of $\mathbb{R}$. Then Y is not compact because the open cover $\mathscr{F} = \left\{\left(\frac{1}{n},1\right] : n \in \mathbb{N}\right\}$ contains no finite subcover of $(0,1]$.

On the other hand, the interval $[0,1]$ is compact.

(g) The open sphere $S_1(0)$ with center at origin and radius 1 is not compact.

Indeed, $S_1(0) \subseteq \bigcup_{n=2}^{\infty} S_{1-\frac{1}{n}}(0)$. However, no finite subcollection of the open cover $\left\{S_{1-\frac{1}{n}}(0) : n = 2,3,\ldots\right\}$ covers $S_1(0)$.

(h) Let (X,d) be a discrete metric space and X be an infinite set. Then X is not compact.

Consider the family $\mathscr{F} = \{\{x\} : x \in X\}$ of subsets of X which is an open cover of X since in the discrete metric space, every subset is an open set and so the singleton set. But there is no finite subcollection of $\mathscr{F}$ that covers X.

Further, any infinite subset of a discrete metric space is not compact.

Lemma 4.1. *Let Y be a subspace of a metric space X. Then Y is compact if and only if every cover of Y by sets open in X contains a finite subcover of Y.*

Proof. Suppose that Y is compact and $\mathscr{F} = \{G_\alpha\}_{\alpha\in\Lambda}$ is an open cover of Y, where each G_α is open in X. Then the collection $\mathscr{F}^* = \{G_\alpha \cap Y : \alpha \in \Lambda\}$ is an open cover of Y by sets $G_\alpha \cap Y$ open in Y. Since Y is compact, $\mathscr{F}^*$ has a finite subcover $\mathscr{C}^*$ of Y, that is, $\mathscr{C}^* = \{G_{\alpha_1} \cap Y, G_{\alpha_2} \cap Y, \ldots, G_{\alpha_n} \cap Y\}$ such that $Y = \bigcup_{k=1}^{n} (G_{\alpha_k} \cap Y)$. This implies that $Y \subseteq \bigcup_{k=1}^{n} G_{\alpha_k}$ and hence $\mathscr{C} = \{G_{\alpha_1}, G_{\alpha_2}, \ldots, G_{\alpha_n}\}$ is a finite subcover of Y.

Conversely, assume that every cover of Y by sets open in X contains a finite subcover of Y. Then we claim that Y is compact. Let $\mathscr{F}^* = \{G_\alpha^*\}_{\alpha\in\Lambda}$ be an open cover of Y, where each G_α^* is open in Y. For each $\alpha \in \Lambda$, choose a set G_α open in X such that $G_\alpha^* = G_\alpha \cap Y$. Since $\mathscr{F}^*$ is an open cover of Y, we have

$$Y = \bigcup_{\alpha\in\Lambda} G_\alpha^* = \bigcup_{\alpha\in\Lambda} (G_\alpha \cap Y) = \left(\bigcup_{\alpha\in\Lambda} G_\alpha\right) \cap Y.$$

This implies that $Y \subseteq \bigcup_{\alpha\in\Lambda} G_\alpha$ and hence $\mathscr{F} = \{G_\alpha\}_{\alpha\in\Lambda}$ is an open cover of Y by sets G_α open in X. By hypothesis, some finite subcollection $\{G_{\alpha_1}, G_{\alpha_2}, \ldots, G_{\alpha_n}\}$ covers Y, that is, $Y \subseteq \bigcup_{k=1}^{n} G_{\alpha_k}$. Thus

$$Y = \left(\bigcup_{k=1}^{n} G_{\alpha_k}\right) \cap Y = \bigcup_{k=1}^{n} (G_{\alpha_k} \cap Y) = \bigcup_{k=1}^{n} G_{\alpha_k}^*.$$

Therefore, $\{G_{\alpha_1}^*, G_{\alpha_2}^*, \ldots, G_{\alpha_n}^*\}$ is a finite subcover of Y and hence Y is compact.

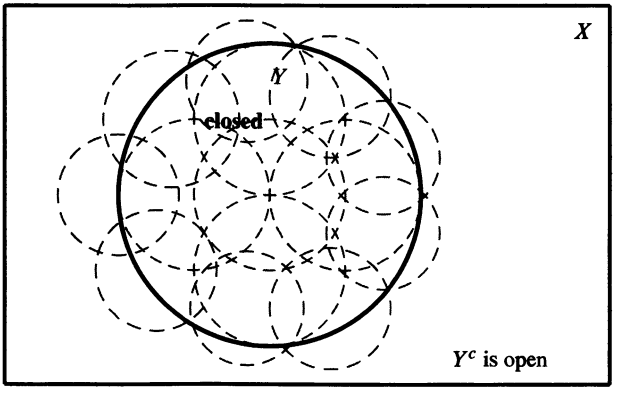
Theorem 4.1. *Every closed subset of a compact metric space is compact.*

Proof. Let Y be a closed subset of a compact metric space X and let $\mathscr{F} = \{G_\alpha\}_{\alpha\in\Lambda}$ be an open cover of Y by sets open in X, that is, $Y \subseteq \bigcup_{\alpha\in\Lambda} G_\alpha$, where each G_α is open in X. Then $\mathscr{F}^* = \mathscr{F} \cup Y^c$ is an open cover of X, because Y^c is open in X. Since X is compact, $\mathscr{F}^*$ has a finite subcover of X, say $\mathscr{C}^*$. If this finite subcover $\mathscr{C}^*$ contains Y^c, that is, if

$$X = G_{\alpha_1} \cup G_{\alpha_2} \cup \cdots \cup G_{\alpha_n} \cup Y^c, \quad \text{where } G_{\alpha_k} \in \mathscr{F},\ k = 1,2,\ldots,n,$$

then discard Y^c since it covers no part of Y; otherwise leave the subcollection alone, that is, $\mathscr{C}^* = \{G_{\alpha_1}, G_{\alpha_2}, \ldots G_{\alpha_n}\}$. Then

$$Y \subseteq G_{\alpha_1} \cup G_{\alpha_2} \cup \cdots \cup G_{\alpha_n}, \quad \text{where } G_{\alpha_k} \in \mathscr{F},\ k = 1,2,\ldots,n.$$



Thus $\mathscr{C}^*$ is a finite subcollection of $\mathscr{F}$ that covers Y. Then by Lemma 4.1, Y is compact.

Remark 4.1. (i) Any intersection of compact sets is compact.
(ii) The finite union of compact sets is compact.

Definition 4.4. A collection $\mathscr{C} = \{C_1, C_2, \ldots\}$ of subsets of a metric space X is said to have the *finite intersection property* if every finite subcollection of $\mathscr{C}$ has nonempty intersection, that is, for every finite collection $\{C_1, C_2, \ldots, C_n\}$ of $\mathscr{C}$, we have $\bigcap_{i=1}^n C_i \neq \emptyset$.

Example 4.3. (a) Consider the following collection of open intervals

$$\mathscr{C} = \left\{(0,1), \left(0,\frac{1}{2}\right), \left(0,\frac{1}{3}\right), \left(0,\frac{1}{4}\right), \ldots\right\}.$$

Then $\mathscr{C}$ has the finite intersection property.

Indeed $(0,a_1) \cap (0,a_2) \cap (0,a_3) \cap \cdots \cap (0,a_m) = (0,b)$, where $b = \min\{a_1, a_2, \ldots, a_m\} > 0$. Therefore, any subclass $\mathscr{C}$ has nonempty intersection even then $\mathscr{C}$ itself has an empty intersection. Thus $\mathscr{C}$ satisfies the finite intersection property.

(b) Consider the following collection of closed infinite intervals in $\mathbb{R}$

$$\mathscr{C} = \{\ldots,(-\infty,-2],(-\infty,-1],(-\infty,0],(-\infty,1],(-\infty,2],\ldots\}.$$

Then $\mathscr{C}$ has an empty intersection, that is, $\bigcap_{n\in\mathbb{Z}} C_n = \emptyset$, where $C_n = (-\infty, n]$. But any finite subclass of $\mathscr{C}$ has a nonempty intersection. So, $\mathscr{C}$ satisfies the finite intersection property.

Theorem 4.2. *A metric space X is compact if and only if every collection of closed sets in X having finite intersection property has nonempty intersection.*

Proof. Let X be a compact metric space and $\mathscr{C} = \{C_\alpha\}_{\alpha\in\Lambda}$ be any collection of closed sets in X having finite intersection property. Then we claim that $\bigcap_{\alpha\in\Lambda} C_\alpha \neq \emptyset$. Assume contrary that $\bigcap_{\alpha\in\Lambda} C_\alpha = \emptyset$. Then by De Morgan's law

$$\bigcup_{\alpha\in\Lambda} (X \setminus C_\alpha) = X \setminus \left(\bigcap_{\alpha\in\Lambda} C_\alpha\right) = X.$$

This implies that $\mathscr{F} = \{X \setminus C_\alpha\}_{\alpha\in\Lambda}$ is an open cover of X since each $X \setminus C_\alpha$ is open. Since X is compact, $\mathscr{F}$ has a finite subcover $\mathscr{C} = \{X \setminus C_{\alpha_i}\}_{i=1}^n$, say. Then $X = \bigcup_{i=1}^n (X \setminus C_{\alpha_i})$. This implies that $X \setminus (\bigcap_{i=1}^n C_{\alpha_i}) = X$ and hence $\bigcap_{i=1}^n C_{\alpha_i} = \emptyset$. This shows that $\mathscr{C}$ does not have the finite intersection property, which is a contradiction of our hypothesis. Hence $\bigcap_{\alpha\in\Lambda} C_\alpha \neq \emptyset$.

Conversely, assume that X is not compact. Then there exists an open cover $\mathscr{F} = \{G_\alpha\}_{\alpha\in\Lambda}$ of X which does not have finite subcover, that is, for every finite subcollection $\mathscr{C} = \{G_{\alpha_1}, G_{\alpha_2}, \ldots, G_{\alpha_n}\}$ of $\mathscr{F}$, we have

$$\bigcup_{i=1}^n G_{\alpha_i} \neq X \quad \Rightarrow \quad X \setminus \left(\bigcup_{i=1}^n G_{\alpha_i}\right) \neq \emptyset \quad \Rightarrow \quad \bigcap_{i=1}^n (X \setminus G_{\alpha_i}) \neq \emptyset.$$

For each $\alpha \in \Lambda$, let $C_\alpha = X \setminus G_\alpha$, then C_α is closed. Since

$$\bigcap_{i=1}^n C_{\alpha_i} = \bigcap_{i=1}^n (X \setminus G_{\alpha_i}) \neq \emptyset,$$

$\{C_\alpha\}_{\alpha\in\Lambda}$ is a collection of closed sets having finite intersection property. By the hypothesis, $\bigcap_{\alpha\in\Lambda} C_\alpha \neq \emptyset$.

On the other hand, since $\mathscr{F} = \{G_\alpha\}_{\alpha\in\Lambda}$ is a cover of X, we have $\bigcup_{\alpha\in\Lambda} G_\alpha = X$ and so

$$\bigcap_{\alpha\in\Lambda} C_\alpha = \bigcap_{\alpha\in\Lambda} (X \setminus G_\alpha) = X \setminus \left(\bigcup_{\alpha\in\Lambda} G_\alpha\right) = \emptyset$$

which is a contradiction. Hence X is compact.

Corollary 4.1. *A metric space X is compact if and only if every collection $\mathscr{C} = \{F_1, F_2, \ldots\}$ of subsets of X having finite intersection property, the intersection $\bigcap_{F\in\mathscr{C}} \overline{F}$ of their closures is nonempty.*

Proof. Let $\{F_1, F_2, \ldots, F_n\}$ be a finite subcollection of $\mathscr{C}$ such that $\bigcap_{i=1}^n F_i \neq \emptyset$. Since for each i, $F_i \subseteq \overline{F}_i$, we have

$$\emptyset \neq \bigcap_{i=1}^{n} F_i \subseteq \bigcap_{i=1}^{n} \overline{F}_i.$$

Therefore, $\bigcap_{i=1}^{n} \overline{F}_i \neq \emptyset$ and hence $\mathscr{C}' = \{\overline{F}_1, \overline{F}_2, \ldots\}$ be a collection of closed sets having finite intersection property. By Theorem 4.2, $\bigcap_{F \in \mathscr{C}} \overline{F} \neq \emptyset$.

Exercise 4.1. (a) Prove that any intersection of compact sets is compact.
(b) Prove that any finite union of compact sets is compact.

Proof. We prove only part (b). Part (a) is left for the readers.

Let $\{Y_k\}_{k=1}^{m}$ be a family of compact sets in a metric space X. Let $\mathscr{F} = \{G_\alpha\}_{\alpha \in \Lambda}$ be an open cover of $\bigcup_{k=1}^{m} Y_k$ by sets open in X. Since each Y_k is compact, there exists finite subset J_k of Λ such that

$$Y_k \subseteq \bigcup_{\alpha_i \in J_k} G_{\alpha_i} \quad \text{for each } k = 1, 2, \ldots m.$$

Therefore, $\bigcup_{k=1}^{m} Y_k \subseteq \bigcup_{\alpha_i \in J} G_{\alpha_i}$ where $J = \bigcup_{k=1}^{m} J_k$ is a finite subset of Λ. Hence every open cover of $\bigcup_{k=1}^{m} Y_k$ by sets open in X has a finite subcover and thus by Lemma 4.1, $\bigcup_{k=1}^{m} Y_k$ is compact.

Exercise 4.2. Prove that a metric space X is compact if and only if every family of closed sets with empty intersection has a finite subfamily with empty intersection.

Proof. Let X be a compact metric space and let $\{C_\alpha\}_{\alpha \in \Lambda}$ be a family of closed sets in X such that $\bigcap_{\alpha \in \Lambda} C_\alpha = \emptyset$. Taking the complements, we have

$$X = [\emptyset]^c = \left[\bigcap_{\alpha \in \Lambda} C_\alpha\right]^c = \bigcup_{\alpha \in \Lambda} C_\alpha^c,$$

that is, $X = \bigcup_{\alpha \in \Lambda} C_\alpha^c$, where each C_α^c is an open set in X since C_α^c is the complement of a closed set C_α. Thus $\mathscr{F} = \{C_\alpha^c\}_{\alpha \in \Lambda}$ is an open cover of X. Since X is compact, there exists a finite subcover $\mathscr{C}$ of cover $\mathscr{F}$, that is, there exist $C_{\alpha_1}^c, C_{\alpha_2}^c, \ldots, C_{\alpha_n}^c \in \mathscr{F}$ such that $\bigcup_{i=1}^{n} C_{\alpha_i}^c = X$. Again, by taking the complement, we have $\bigcap_{i=1}^{n} C_{\alpha_i} = \emptyset$.

Conversely, assume that the hypothesis holds. Then we claim that X is compact. Let $\mathscr{F} = \{G_\alpha\}_{\alpha \in \Lambda}$ be an open cover of X. Then $\bigcup_{\alpha \in \Lambda} G_\alpha = X$. Taking the complements, we have

$$\emptyset = [X]^c = \left[\bigcup_{\alpha \in \Lambda} G_\alpha\right]^c = \bigcap_{\alpha \in \Lambda} G_\alpha^c,$$

that is, $\bigcap_{\alpha \in \Lambda} G_\alpha^c = \emptyset$. Thus, $\{G_\alpha^c\}_{\alpha \in \Lambda}$ is a family of closed sets with empty intersection. So, by the given hypothesis, there exists a finite subfamily $\{G_{\alpha_1}^c, G_{\alpha_2}^c, \ldots, G_{\alpha_n}^c\}$ such that $\bigcap_{i=1}^{n} G_{\alpha_i}^c = \emptyset$. Taking the complements, we have

$$X = [\emptyset]^c = \left[\bigcap_{i=1}^{n} G_{\alpha_i}^c\right]^c = \bigcup_{i=1}^{n} G_{\alpha_i}.$$

Thus $\{G_{\alpha_1}, G_{\alpha_2}, \ldots, G_{\alpha_n}\}$ is a finite subcover of $\mathscr{F}$ and hence X is compact.

Exercise 4.3. Let $\{C_n\}_{n=1}^{\infty}$ be a sequence of nonempty closed subsets of a compact metric space X such that $C_{n+1} \subset C_n$ for each $n \in \mathbb{N}$. Prove that $\bigcap_{n=1}^{\infty} C_n \neq \emptyset$.

Proof. Since $C_1 \supset C_2 \supset C_3 \supset \cdots \supset C_n \supset C_{n+1} \supset \cdots$ is a nested sequence of nonempty closed sets in a compact metric space X. Then the collection $\mathscr{C} = \{C_n\}_{n=1}^{\infty}$ satisfies the finite intersection property (prove it). By Theorem 4.2, we have $\bigcap_{n=1}^{\infty} C_n \neq \emptyset$.

4.2 Sequentially Compact Spaces

In the real line $\mathbb{R}$, we have the following well known result.

Theorem 4.3 (Bolzano-Weierstrass Theorem). *If A is a closed and bounded subset of $\mathbb{R}$, then every infinite subset of A has a limit point in A.*

This result is no longer valid in the setting of general metric spaces. For example, consider the metric space (X,d), where $X = (0,1]$ and d is the usual metric. Then X itself is closed and bounded. Let $A = \left\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\right\}$ be an infinite subset of X. We observe that A has exactly one limit point $0 \notin X$. This motivates the following concept.

Definition 4.5. A metric space X is said to have the *Bolzano-Weierstrass property* if every infinite subset of X has a limit point.

Example 4.4. By Bolzano-Weierstrass Theorem 4.3, every closed and bounded interval $A = [a,b]$ has the Bolzano-Weierstrass property.

Theorem 4.4. *Every compact metric space has the Bolzano-Weierstrass property.*

Proof. Let A be an infinite subset of a compact metric space X. Assume that A has no limit point. Then each point x of X is not a limit point of A, and so, there exists $r_x > 0$ such that $S_{r_x}(x)$ does not contain any point of A other than x. Therefore, $S_{r_x}(x) \cap A$ contains no points of A other than x, that is,

$$S_{r_x}(x) \cap A = \begin{cases} \emptyset & \text{if } x \notin A \\ \{x\} & \text{if } x \in A. \end{cases} \tag{4.1}$$

Since at each $x \in X$, we have an open sphere $S_{r_x}(x)$ which contains no point of A other than its center, the family $\mathscr{C} = \{S_{r_x}(x) : x \in X\}$ of these open spheres forms an open cover of X. Since X is compact, there exists a finite subcover of $\mathscr{C}$, say

$$\left\{S_{r_{x_i}}(x_i) : x_i \in X \text{ and } i = 1,2,\ldots,n\right\}$$

such that $X = \bigcup_{i=1}^{n} S_{r_{x_i}}(x_i)$. Then

$$\begin{aligned} A = A \cap X &= A \cap \left(\bigcup_{i=1}^{n} S_{r_{x_i}}(x_i)\right) \\ &= \bigcup_{i=1}^{n} \left(A \cap S_{r_{x_i}}(x_i)\right) \\ &\subseteq \bigcup_{i=1}^{n} \{x_i\} \quad \text{by (4.1)} \\ &= \{x_1, x_2, \ldots x_n\}. \end{aligned}$$

Consequently, A is a finite set which is a contradiction of our assumption. Hence A has a limit point.

Definition 4.6. A metric space X is said to be *sequentially compact* if every sequence in X has a convergent subsequence.

A subset A of a metric space X is said to be *sequentially compact* if every sequence in A contains a subsequence which converges to a point in A.

Example 4.5. (a) Every finite subset of a metric space X is sequentially compact.

Indeed, if A is a finite subset of a metric space X and $\{x_1, x_2, \ldots, x_n, \ldots\}$ is a sequence in A, then at least one of the elements in A, say x_{n_0} must appear in infinite number of times in the sequence. Hence $\{x_{n_0}, x_{n_0}, \ldots\}$ is a subsequence of the sequence $\{x_1, x_2, \ldots, x_n, \ldots\}$ that converges to the point $x_{n_0} \in A$.

(b) The open interval $A = (0,1)$ on the real line $\mathbb{R}$ with the usual metric is not sequentially compact.

Consider the sequence $\{x_n\} = \{\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\}$ in A. Then $x_n \to 0$, and therefore, every subsequence will also converges to 0. But $0 \notin A$. In other words, the sequence $\{x_n\}$ in A does not contain a subsequence that converges to a point in A, that is, A is not sequentially compact.

In order to prove the following relation

$$\text{compactness} \Leftrightarrow \text{Bolzano-Weierstrass property} \Leftrightarrow \text{sequentially compactness}$$

we first establish the equivalence between sequentially compactness and Bolzano-Weierstrass property.

Theorem 4.5. *A metric space X is sequentially compact if and only if it has the Bolzano-Weierstrass property.*

Proof. Let X be a sequentially compact metric space. Then we claim that every infinite subset of X has a limit point. Let A be an infinite subset of X. Then we can extract a sequence $\{x_n\}$ with infinitely many distinct points from A. Since X is sequentially compact, the sequence $\{x_n\}$ has a convergent subsequence, say, $\{x_{n_k}\}$. Let $x_{n_k} \to x$ in X. Then we prove that x is a limit point of A.

Let $\varepsilon > 0$ be given. Since $x_{n_k} \to x$ as $k \to \infty$, there exists a positive integer N such that

$$x_{n_k} \in S_\varepsilon(x) \quad \text{for all } k > N.$$

Since $\{x_n\}$ consists infinitely many distinct points from A, we have $x_{n_k} \neq x$ for all $k > N$. Thus $S_\varepsilon(x)$ contains the element x_{n_k}, for all $k > N$, of A different from x. Hence x is a limit point of A. Thus every infinite subset of X has a limit point and hence X has the Bolzano-Weierstrass property.

Conversely, assume that X has the Bolzano-Weierstrass property. Then every infinite subset of X has a limit point. Let $\{x_n\}$ be an arbitrary sequence in X. Then we have to prove that $\{x_n\}$ has a convergent subsequence. Let A be the range set of the sequence $\{x_n\}$, that is, $A = \{x_1, x_2, x_3, \ldots, x_n, \ldots\}$ is a set of points $x_1, x_2, x_3, \ldots, x_n, \ldots$. Then there are two cases:

CASE 1. If A is a finite set, then the sequence $\{x_n\}$ has a point x which appears infinitely many times in the sequence $\{x_n\}$ (that is, $\{x_n\} = \{x_1, x_2, \ldots, x_n, x, x, \ldots\}$).

Then we have a constant subsequence $\{x,x,x,\ldots,x\ldots\}$ of $\{x_n\}$ which certainly converges to x.

CASE 2. If A is an infinite set, then the sequence $\{x_n\}$ has infinitely many distinct points. By our assumption, the set A has a limit point, say, $x \in X$. Then $S_1(x)$ must have infinitely many points from A. Let us choose one such element, say, x_{n_1} other than x. Again $S_{1/2}(x)$ must have infinitely many points from A and in the same way, we get x_{n_2} other than x, $n_2 > n_1$. Similarly, $S_{1/3}(x)$ must have infinitely many points from A and we get $x_{n_3} \neq x$, $n_3 > n_2$, and so on. In this way, we obtain a subsequence $\{x_{n_k}\}$ of the sequence $\{x_n\}$ such that

$$x_{n_k} \in S_{1/k}(x) \quad \text{for all } k = 1,2,\ldots,$$

that is,

$$d(x_{n_k},x) < \frac{1}{k} \to 0 \quad \text{as } k \to \infty$$

and hence $x_{n_k} \to x$. Thus the sequence $\{x_n\}$ has a convergent subsequence $\{x_{n_k}\}$, and therefore, X is sequentially compact.

Corollary 4.2. *Every compact metric space is sequentially compact.*

Proof. It follows from Theorems 4.5 and 4.4.

Theorem 4.6 (Heine-Borel Theorem). *A subset of* $\mathbb{R}$ *is closed and bounded if and only if it is compact.*

As an application of Corollary 4.2, we present a general form of the Heine-Borel theorem in the setting of metric spaces.

Theorem 4.7. *Every compact subset of a metric space is closed and bounded.*

Proof. Let (X,d) be a metric space and A be a compact subset of X. If A is finite, then there is nothing to prove because every finite set is closed and bounded. So, we assume that A is an infinite set.

Let $x \in \overline{A}$ be arbitrary. Then by Theorem 2.3, there exists a sequence $\{x_n\} \subset A$ such that $x_n \to x$. Since A is compact, by Corollary 4.2, it is sequentially compact, and therefore, $\{x_n\}$ has a subsequence that converges in A. Let $\{x_{n_k}\}$ be a convergent subsequence of $\{x_n\}$. Then $x_{n_k} \to x \in A$ and hence $\overline{A} \subset A$. Therefore, $\overline{A} = A$ and thus A is a closed set.

Further, let A be unbounded. Then we can find pairs of points of A at arbitrary large distances apart, that is, there would exist x and y in A such that $d(x,y) > M$ for some given $M > 0$. Consider the family $\mathscr{F} = \{S_1(x) : x \in A\}$ of open spheres with centered at a point of A and radius 1. Then clearly,

$$A \subseteq \bigcup_{x \in A} S_1(x).$$

This implies that $\mathscr{F}$ is an open cover of A. Since A is compact, $\mathscr{F}$ has a finite subcover $\mathscr{C} = \{S_1(x_i) : i = 1,2,\ldots,n\}$, say. Then

$$A \subseteq \bigcup_{i=1}^{n} S_1(x_i).$$

Let $k = \max\{d(x_i,x_j) : i,j = 1,2,\ldots,n,\ i \neq j\}$. Also, let $M = k+2$, then

$$d(x,y) > k+2. \tag{4.1}$$

Since x and y are in A, there exist x_i and x_j such that $x \in S_1(x_i)$ and $y \in S_1(x_j)$. So,

$$d(x,y) \leq d(x,y_i) + d(x_i,x_j) + d(x_j,y) < k+2$$

which contradicts (4.1). Hence A is bounded.

Remark 4.2. The converse of the above theorem need not be true. For example, consider an infinite subset A of a discrete metric space X. Then A is closed and bounded because the open sphere $S_{\frac{1}{2}}(x)$ is the set $\{x\}$ alone and $d(x,y) \leq 1$ for all $x,y \in A$. But the open cover $\{\{x\} : x \in A\}$ of A has no finite subcover. Thus A is not compact.

In fact, a subset of $\mathbb{R}^n$ with the usual metric is compact if and only if it is closed and bounded.

4.3 Totally Bounded Spaces

Definition 4.7. Let (X,d) be a metric space and $\varepsilon > 0$ be given. A subset A of X is called an *ε-net* if A is finite and $X = \bigcup_{x \in A} S_\varepsilon(x)$, that is, if A is finite and its points are scattered through X in such a way that each point of X is distant by less than ε from at least one point of A.

In other words, a finite subset $A = \{x_1,x_2,\ldots,x_n\}$ of X is an ε-net for X if for every point $y \in X$, there exists an $x_{i_0} \in A$ such that $d(y,x_{i_0}) < \varepsilon$.

Example 4.6. Let $X = \{(x,y) \in \mathbb{R} \times \mathbb{R} : x^2 + y^2 < 4\}$, that is, X is the open sphere centered at the origin and radius 2. If $\varepsilon = \frac{3}{2}$, then the set

$$A = \{(1,-1),(1,0),(1,1),(0,-1),(0,0),(0,1),(-1,-1),(-1,0),(-1,1)\}$$

is an ε-net for X.

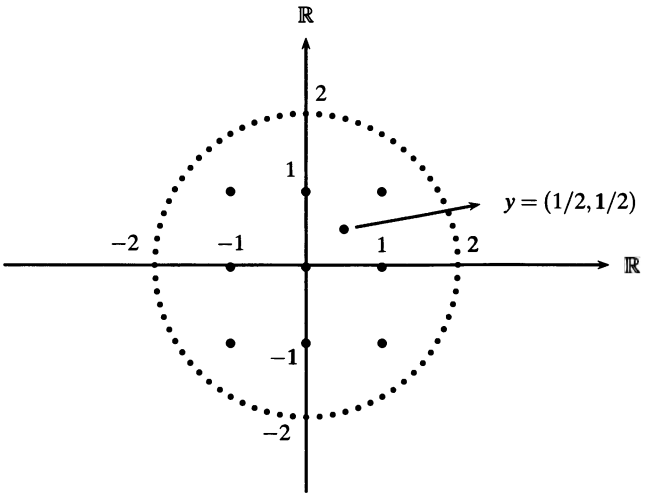
On the other hand, if $\varepsilon = 1/2$, then A is not an ε-net for X. For example, the point $y = \left(\frac{1}{2},\frac{1}{2}\right)$ belongs to X but the distance between y and any point in A is greater than $\frac{1}{2}$.

Definition 4.8. A metric space (X,d) is said to be *totally bounded* if it has an ε-net for each $\varepsilon > 0$.

Remark 4.3. Every totally bounded metric space X is bounded but converse is not true in general.

Since X is totally bounded, it has an ε-net $A = \{x_1,x_2,\ldots,x_n\}$ for each $\varepsilon > 0$. Then $X = \bigcup_{i=1}^n S_\varepsilon(x_i)$. Since finite union of bounded sets is bounded, it follows that X is bounded.

Example 4.7. Under the usual metric $d(x,y) = |x-y|$, the real line $\mathbb{R}$ is neither bounded nor totally bounded. Under the metric $d^*(x,y) = \min\{|x-y|,1\}$, the real line $\mathbb{R}$ is bounded but not totally bounded.



Theorem 4.8. *Every sequentially compact metric space is totally bounded and complete.*

Proof. Let (X,d) be a sequentially compact metric space. Assume to the contrary that X is not totally bounded. Then there exists an $\varepsilon > 0$ such that there does not exist any ε-net. Choose a point $x_1 \in X$ and form the open sphere $S_\varepsilon(x_1)$. Then, clearly $S_\varepsilon(x_1) \neq X$ for otherwise $\{x_1\}$ would be an ε-net for X. Let $x_2 \in X$ be an element such that $x_2 \notin S_\varepsilon(x_1)$, that is, $d(x_1,x_2) \geq \varepsilon$. Form the open sphere $S_\varepsilon(x_2)$ and the set $S_\varepsilon(x_1) \cup S_\varepsilon(x_2)$. Then $S_\varepsilon(x_1) \cup S_\varepsilon(x_2) \neq X$ since in other case $\{x_1,x_2\}$ would be an ε-net for X. Let $x_3 \in X$ be such that $x_3 \notin S_\varepsilon(x_1) \cup S_\varepsilon(x_2)$, that is,

$$d(x_1,x_3) \geq \varepsilon \quad \text{and} \quad d(x_2,x_3) \geq \varepsilon.$$

Form an open sphere $S_\varepsilon(x_3)$ and a set $S_\varepsilon(x_1) \cup S_\varepsilon(x_2) \cup S_\varepsilon(x_3)$. Then $S_\varepsilon(x_1) \cup S_\varepsilon(x_2) \cup S_\varepsilon(x_3) \neq X$, otherwise the set $\{x_1,x_2,x_3\}$ would be an ε-net for X.

Continuing in this way, we obtain a sequence $\{x_n\}$ of elements of X such that

$$x_n \notin \bigcup_{i=1}^{n-1} S_\varepsilon(x_i) \quad \text{for } n = 2,3,\ldots,$$

that is,

$$d(x_i,x_n) \geq \varepsilon \quad \text{for } i = 1,2,\ldots,n-1 \text{ and } n = 2,3,\ldots \text{ with } n \neq i.$$

Consequently,

$$d(x_n,x_m) \geq \varepsilon \quad \text{for all } n,m \text{ and } n \neq m.$$

Thus $\{x_n\}$ is a sequence with no Cauchy subsequence and hence $\{x_n\}$ cannot have any convergent subsequence, contradicting our hypothesis that X is sequentially compact. Hence X is totally bounded.

Now, we prove that X is complete. For that, we have to show that every Cauchy sequence converges to a point of X.

Let $\{x_n\}$ be a Cauchy sequence. Then for each $\varepsilon > 0$, there exists a positive integer N such that

$$d(x_m, x_n) < \varepsilon \quad \text{whenever} \quad m > n > N.$$

Since X is sequentially compact, every sequence, in particular, a Cauchy sequence, has a convergent subsequence. Therefore, this Cauchy sequence $\{x_n\}$ contains a subsequence $\{x_{k_n}\}$, say, which converges to a point $x \in X$, that is,

$$\lim_{n\to\infty} d(x_{k_n}, x) = 0.$$

Since $\{k_n\}$ is an increasing sequence of integers, $k_m \geq m$. We now have

$$0 \leq d(x_n, x) \leq d(x_n, x_{k_m}) + d(x_{k_m}, x) < \varepsilon + d(x_{k_m}, x)$$

whenever $m > n > N$. Taking $m \to \infty$, then we have

$$0 \leq d(x_n, x) \leq \varepsilon \quad \text{whenever } n \geq N.$$

Therefore, the Cauchy sequence $\{x_n\}$ converges to $x \in X$.

Remark 4.4. It is not necessary that a totally bounded metric space is sequentially compact. For example, consider the metric subspace $A = (-1, 1)$ of $\mathbb{R}$ under the usual metric $d(x, y) = |x - y|$ for all $x, y \in A$. Then $A = (-1, 1)$ is totally bounded but not sequentially compact. But every totally bounded complete metric space is sequentially compact.

Theorem 4.9. *Every totally bounded complete metric space is sequentially compact.*

Proof. Let (X, d) be a totally bounded and complete metric space. Let $\{x_n\}$ be any sequence in X. Since X is totally bounded, it can be covered by a finite number of open spheres of radius 1, and at least one of these open spheres will contain an infinite number of members of the sequence $\{x_n\}$. Let N_1 be a such open sphere and let $\{x_{n_1}\}$ be a subsequence of the sequence $\{x_n\}$ in N_1.

Since X is covered by a finite number of open spheres and N_1 is contained in X, therefore, N_1 can also be covered by a finite number of open spheres of radius $\frac{1}{2}$. By the same argument, there would be an open sphere N_2 of radius $\frac{1}{2}$ which contains a subsequence $\{x_{n_2}\}$ of $\{x_{n_1}\}$. Since $N_2 \subset N_1$, it follows that $x_{n_2} \in N_1$. Continuing in this way, for any positive integer k, we obtain an open sphere N_k of radius $\frac{1}{k}$ which is contained in N_{k-1} and contains a subsequence $\{x_{n_k}\}$ of $\{x_{n_{k-1}}\}$, that is, the diameter of N_k, $\delta(N_k) < \frac{2}{k}$ and $x_{n_k} \in N_k$, where $n_k > n_{k-1}$. Since $x_{n_k}, x_{n_{k+1}}, \ldots$ all lie in N_k and the diameter $\delta(N_k) < \frac{2}{k}$, it follows that

$$d(x_{n_k}, x_{n_{k+m}}) < \frac{2}{k}.$$

For any $\varepsilon > 0$, choose $k > \frac{2}{\varepsilon}$. Then

$$d(x_{n_k}, x_{n_{k+m}}) < \varepsilon$$

and hence $\{x_{n_k}\}$ is a Cauchy sequence of the sequence $\{x_n\}$. Since X is complete, this subsequence converges to a point of X. Hence X is sequentially compact.

Corollary 4.3. *A closed subspace of a complete metric space is sequentially compact if and only if it is totally bounded.*

Theorem 4.10. *Every totally bounded metric space is separable.*

Proof. Let (X,d) be a totally bounded metric space. Then for each $n \in \mathbb{N}$, X has a $\frac{1}{n}$-net which is denoted by A_n. Since each A_n is a finite set, the set $A = \bigcup_{n=1}^{\infty} A_n$ is countable. We further prove that $\overline{A} = X$, that is, if $x \in X$ then we claim that $x \in \overline{A}$.

Let $x \in X$ and $r > 0$ be given. Choose a positive integer n such that $\frac{1}{n} < r$. Let $A_n = \{a_1, a_2, \ldots, a_k\}$ since each A_n is finite. Then

$$X = \bigcup_{j=1}^{k} S_{\frac{1}{n}}(a_j)$$

and so $x \in S_{\frac{1}{n}}(a_j)$ for some $j = 1,2,\ldots,k$. Therefore,

$$d(x,a_j) < \frac{1}{n} < r$$

and thus $a_j \in S_r(x)$, that is, $S_r(x) \cap A \neq \emptyset$ because $a_j \in A_n \subset A$. This implies that $x \in \overline{A}$ and hence $\overline{A} = X$.

Definition 4.9. Let $\mathscr{F} = \{G_\alpha\}_{\alpha \in \Lambda}$ be an open cover of a metric space X. A real number $\lambda > 0$ is called a *Lebesgue number* for the given open cover $\mathscr{F}$ if for each subset of X whose diameter is less than λ is contained in at least one G_α, that is, if every subset $A \subset X$ with $\delta(A) < \lambda$ then $A \subset G_\alpha$ for at least one $\alpha \in \Lambda$.

Theorem 4.11 (Lebesgue's Covering Lemma). *In a sequentially compact metric space, every open cover has a Lebesgue number.*

Proof. Let (X,d) be a sequentially compact metric space and let $\mathscr{F} = \{G_\alpha\}_{\alpha \in \Lambda}$ be an open cover of X. Assume that $\mathscr{F}$ does not have any Lebesgue number. Then for each $n \in \mathbb{N}$, there exists a subset B_n of X such that

$$0 < \delta(B_n) < \frac{1}{n} \quad \text{and} \quad B_n \not\subset G_\alpha \quad \text{for all } \alpha \in \Lambda. \tag{4.2}$$

For each $n \in \mathbb{N}$, choose a point $x_n \in B_n$. Then $\{x_n\}$ is a sequence of points of X. Since X is sequentially compact, $\{x_n\}$ has a subsequence $\{x_{n_k}\}$ converges to a point $x \in X$.

Since $x \in X$ and $\mathscr{F}$ is an open cover of X, $x \in G_{\alpha_0}$ for some $\alpha_0 \in \Lambda$. But G_{α_0} being an open set containing x, there exists an open sphere $S_\varepsilon(x)$ such that $x \in S_\varepsilon(x) \subseteq G_{\alpha_0}$. Since $\{x_{n_k}\}$ converges to x, there exists a positive integer N such that for all $k \geq N$, we have

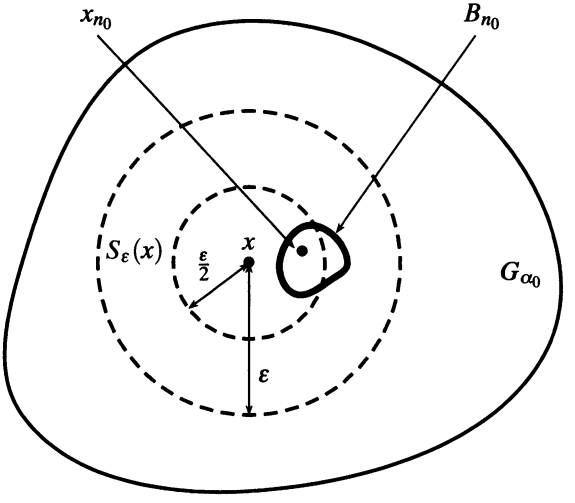
$$d(x_{n_k}, x) < \frac{\varepsilon}{2}.$$

Let $n_0 \geq N$ such that $\frac{1}{n_0} < \frac{\varepsilon}{2}$. Then from above inequality, we have

$$d(x_{n_k}, x) < \frac{\varepsilon}{2}, \quad \text{that is,} \quad x_{n_k} \in S_{\frac{\varepsilon}{2}}(x) \quad \text{for all } n_k \geq n_0.$$

In particular, $x_{n_0} \in S_{\frac{\varepsilon}{2}}(x)$. In view of (4.2), we have

$$x_{n_0} \in B_{n_0} \quad \text{and} \quad 0 < \delta\left(B_{n_0}\right) < \frac{1}{n_0} < \frac{\varepsilon}{2}.$$



Since x_{n_0} belongs to B_{n_0} as well as to $S_{\frac{\varepsilon}{2}}(x)$, $B_{n_0} \cap S_{\frac{\varepsilon}{2}}(x) \neq \emptyset$. This implies that B_{n_0} will also intersect with $S_\varepsilon(x)$. Since $\delta\left(B_{n_0}\right) < \frac{\varepsilon}{2}$ and B_{n_0} intersects with $S_\varepsilon(x)$, we have

$$B_{n_0} \subset S_\varepsilon(x) \subseteq G_{\alpha_0}.$$

But this contradicts with the fact that $B_{n_0} \not\subset G_\alpha$ for all $\alpha \in \Lambda$. Hence $\mathscr{F}$ has a Lebesgue number.

Theorem 4.12. *Every sequentially compact metric space is compact.*

Proof. Let X be a sequentially compact metric space. Then by Theorem 4.8, it is totally bounded. Then for a given $\varepsilon > 0$, X has an ε-net A, say. Then $A = \{x_1, x_2, \ldots, x_n\}$ is finite subset of X and

$$X = \bigcup_{k=1}^{n} S_\varepsilon(x_k).$$

Let $\mathscr{F} = \{G_\alpha\}_{\alpha \in \Lambda}$ be an open cover of X. Since X is sequentially compact, by Lebesgue's covering lemma, this open cover $\mathscr{F}$ has a Lebesgue number λ, say. We put $\varepsilon = \frac{\lambda}{3}$. For each $k = 1, 2, \ldots, n$, we have

$$\delta\left(S_\varepsilon(x_k)\right) < 2\varepsilon = \frac{2\lambda}{3} < \lambda.$$

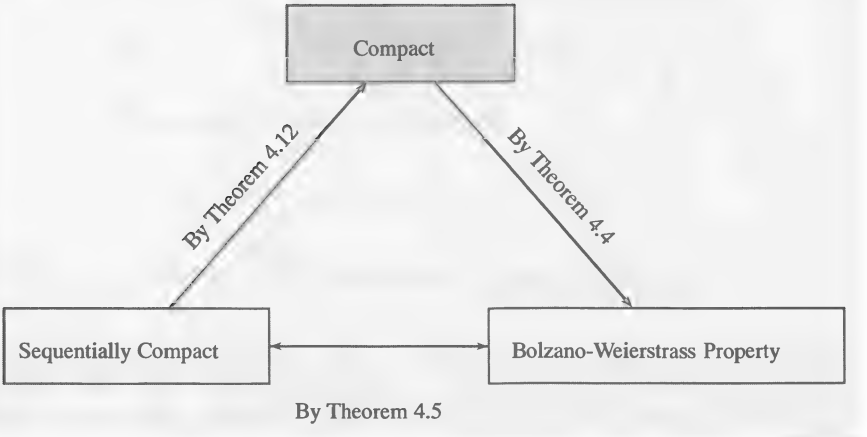
By the definition of a Lebesgue number, for each $k = 1, 2, \ldots, n$, we can find an open set $G_{\alpha_k} \in \mathscr{F}$ such that $S_\varepsilon(x_k) \subseteq G_{\alpha_k}$. This implies that

$$\bigcup_{k=1}^{n} S_\varepsilon(x_k) \subseteq \bigcup_{k=1}^{n} G_{\alpha_k}.$$

Since $X = \bigcup_{k=1}^{n} S_\varepsilon(x_k)$, we have $X = \bigcup_{k=1}^{n} G_{\alpha_k}$. Hence the class $\{G_{\alpha_1}, G_{\alpha_2}, \ldots, G_{\alpha_n}\}$ is a finite subcover of X, and therefore, X is compact.

Corollary 4.4. *A closed subspace of a complete metric space is compact if and only if it is totally bounded.*

Finally, we have the following relationships.



Exercise 4.4. Let A be a compact subset of a metric space X. Show that the derived set A' of A is compact.

Exercise 4.5. Which of the following sets are compact? Justify your answer.

(a) $A = [0,2) \subset \mathbb{R}$.
(b) $A = [0,1] \cup [2,3] \subset \mathbb{R}$.
(c) $A = \{x \in \mathbb{R} : x \geq 0\} \subset \mathbb{R}$.
(d) $A = \mathbb{Q} \cap [0,1] = \{x \in \mathbb{R} : 0 \leq x \leq 1 \text{ and } x \text{ is irrational}\}$.
(e) $A = \{1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots\} \cup \{0\}$.
(f) $A = \{(x,y) \in \mathbb{R}^2 : y = 0\}$.
(g) $A = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 4\} \cup \{(1,2)\}$.
(h) $A = \{(x,y) \in \mathbb{R}^2 : 2x^2 + y^2 = 1\}$.
(i) $A = \{(x,y) \in \mathbb{R}^2 : 0 \leq x \leq 1,\ 0 \leq y \leq 1\}$.

Solution. (b) The set $A = [0,1] \cup [2,3] \subset \mathbb{R}$ is compact because it is closed and bounded (Heine-Borel Theorem).

(c) The set $A = \{x \in \mathbb{R} : x \geq 0\} \subset \mathbb{R}$ is not compact because it is unbounded.

(d) The set $A = \mathbb{Q} \cap [0,1] = \{x \in \mathbb{R} : 0 \leq x \leq 1 \text{ and } x \text{ is irrational}\}$ is not compact because $x = \frac{1}{4} \notin A$, but every interval around it contains irrational numbers which are in A. Therefore, A^c is not open so A is not closed. Therefore, A can not be compact.

(e) Let $\{G_\alpha\}$ be an arbitrary open cover of A. Then the point 0 lies in one of these open sets. Assume that $0 \in G_{\alpha_0}$ for some α_0. Since G_{α_0} is open and $\frac{1}{n} \to 0$, there is an N such that $\frac{1}{N}, \frac{1}{N+1}, \ldots$ all lie in G_{α_0}. Since $\{G_\alpha\}$ is an open cover of A, there must exist $G_{\alpha_1}, G_{\alpha_2}, \ldots, G_{\alpha_N}$ (not necessarily all distinct from one another, but we can rename and give multiple names to the sets as is needed) such that $1, 1/2, \ldots, 1/N \in G_{\alpha_1} \cup G_{\alpha_2} \cup \cdots \cup G_{\alpha_N}$. Then $\{G_{\alpha_0}, G_{\alpha_1}, G_{\alpha_2}, \ldots, G_{\alpha_N}\}$ is a finite subcover of A. Therefore, every open cover has a finite subcover and so A is compact.

Exercise 4.6. Let A and B be subsets of a metric space X such that A is closed and B is compact. Prove that $A \cap B$ is compact.

Proof. Let $\{G_\alpha\}_{\alpha \in \Lambda}$ be an open cover of $A \cap B$ by sets open in X. Let $G_\beta = X \setminus A$. Then G_β is open in X since A is closed. Let us denote by $J = \Lambda \cup \{\beta\}$. We claim that $\{G_\alpha\}_{\alpha \in J}$ is an open cover of B. For that, let $b \in B$. Then either $b \in G_\beta$ (if $b \notin A$), or else $b \in A \cap B \subseteq \bigcup_{\alpha \in \Lambda} G_\alpha$. This implies that $b \in G_\alpha$ for some $\alpha \in \Lambda$. In either case $b \in G_\alpha$ for some $\alpha \in J$. Therefore, $B \subseteq \bigcup_{\alpha \in J} G_\alpha$ and so $\{G_\alpha\}_{\alpha \in J}$ is an open cover of B. Since B is compact, this open cover has a finite subcover, that is, there exists a finite subset L of J such that $B \subseteq \bigcup_{\alpha_i \in L} G_{\alpha_i}$. The set $L^\beta = L - \{\beta\}$ is a finite subset of $J - \{\beta\} = \Lambda$. Now we show that $A \cap B \subseteq \bigcup_{\alpha_i \in L^\beta} G_{\alpha_i}$. For this, suppose that $b \in A \cap B$. Then $b \in B \subseteq \bigcup_{\alpha_i \in L} G_{\alpha_i}$ and so $b \in G_{\alpha_i}$ for some $\alpha_i \in L$. But $b \in A$ and so $b \notin X \setminus A = G_\beta$. Thus $b \in G_{\alpha_i}$ for some $\alpha_i \in L^\beta$. Thus the open cover $\{G_\alpha\}_{\alpha \in \Lambda}$ of $A \cap B$ has a finite subcover, namely $\{G_{\alpha_i}\}_{\alpha_i \in L^\beta}$. Hence $A \cap B$ is compact.

Exercise 4.7. Let $\{A_\alpha\}_{\alpha \in \Lambda}$ be any family of compact subsets of a metric space X. Show that $B = \bigcap_{\alpha \in \Lambda} A_\alpha$ is compact, while $\bigcup_{\alpha \in \Lambda} A_\alpha$ is not necessarily compact.

Proof. Choose a fixed $\alpha_0 \in \Lambda$ and put $A = \bigcap_{\alpha \neq \alpha_0} A_\alpha$ and take $B = A_{\alpha_0}$. Since every compact set is closed, each A_α is closed and so A is closed being an intersection of closed sets. Since B is compact, by Exercise 4.6, $A \cap B$ is compact and so is $A \cap B = \bigcap_{\alpha \in \Lambda} A_\alpha$.

Arbitrary union of compact sets need not be compact. For example, consider the sets $A_n = [-n, n]$ for all positive integer n, in the usual metric space $\mathbb{R}$. Then each A_n is compact since it is closed and bounded. But $\bigcup_{n=1}^{\infty} A_n = \mathbb{R}$ which is not compact.

Exercise 4.8. Given an example of a metric space X (other than $\mathbb{R}$) having Bolzano-Weierstrass property.

Exercise 4.9. Show that no finite subcollection of the open cover $\mathscr{F} = \{S_{1-\frac{1}{n}}(0) : n = 2, 3, \ldots\}$ of $S_1(0)$ covers $S_1(0)$.

Exercise 4.10. Prove that the real line $\mathbb{R}$ under the metric $d^*(x, y) = \min\{|x - y|, 1\}$ for all $x, y \in \mathbb{R}$ is bounded but not totally bounded.

Exercise 4.11. Let $\mathbb{R}$ be the usual metric space and $Y = (0, 1]$ be a subspace of $\mathbb{R}$. Prove that the open cover $\mathscr{F} = \left\{\left(\frac{1}{n}, 1\right] : n \in \mathbb{N}\right\}$ of Y contains no finite subcover.

Exercise 4.12. Prove that a subset A of the usual metric space $\mathbb{R}^n$ is compact if and only if it is closed and bounded.

Proof. Let A be a compact subset of $\mathbb{R}^n$. Then by Theorem 4.7, it is closed and bounded.

Conversely, assume that A is closed and bounded. Fix $\varepsilon > 0$. The set

$$\tilde{C} = \left\{ \left(\frac{\varepsilon k_1}{n}, \frac{\varepsilon k_2}{n}, \ldots, \frac{\varepsilon k_n}{n} \right) : k_i \in \mathbb{Z}, 1 \leq i \leq n \right\}$$

is an ε-net for $\mathbb{R}^n$. Since A is bounded, there exists $r > 0$ such that $A \subset S_r(0)$. Then the set

$$C = \tilde{C} \cap S_r(0)$$

is a finite ε-net for A. So, A is totally bounded. Then by Theorem 4.9, A is sequentially compact and Theorem 4.12 implies that A is compact.

Exercise 4.13. Let X be a metric space and $\{x_n\}$ be a sequence in X such that $x_n \to x$ as $n \to \infty$. Show that the subset $\{x\} \cup \{x_n : n \in \mathbb{N}\}$ of X is compact.

Proof. Let $\mathscr{F}$ be an open cover for $\{x\} \cup \{x_n : n \in \mathbb{N}\}$. Then there exists an open set $G \in \mathscr{F}$ such that $x \in G$. Since $x_n \to x$ and G is open, there exists a positive integer N such that G contains all points of the set $\{x_n : n \in \mathbb{N}\}$ for all $n > N$. For each $i = 1, 2, \ldots, N$, let $G_i \in \mathscr{F}$ such that $x_i \in G_i$. Then $G \cup G_1 \cup \cdots \cup G_N$ is finite subcover of the cover $\mathscr{F}$. Hence the set $\{x\} \cup \{x_n : n \in \mathbb{N}\}$ of X is compact.

Exercise 4.14. Let A be a compact subset of a metric space (X, d). Show that for any subset B of X, there exists a point $x \in A$ such that $\rho(x, B) = \rho(A, B)$.

Proof. Let $\rho(A, B) = \varepsilon$. Then $\varepsilon = \rho(A, B) = \inf\{d(x, y) : x \in A, y \in B\} \leq d(x, y)$ for all $x \in A$ and $y \in B$. Therefore, for any positive integer N, there exist $x_n \in A$, $y_n \in B$ such that

$$\varepsilon \leq d(x_n, y_n) < \varepsilon + \frac{1}{n} \quad \text{for all } n > N.$$

Since A is compact, it is sequentially compact, and therefore, the sequence $\{x_n\} \subseteq A$ has a subsequence $\{x_{n_k}\}$ converges to a point $x \in A$. We claim that $\rho(x, B) = \rho(A, B) = \varepsilon$. Suppose contrary that $\rho(x, B) > \varepsilon$. Then for some $\delta > 0$, $\varepsilon + \delta = \rho(x, B) = \inf_{y \in B} d(x, y)$ and so $d(x, y_n) \geq \varepsilon + \delta$ for all $y_n \in B$. Since $x_{n_k} \to x$ as $k \to \infty$, there exists a positive integer N such that

$$d(x_{n_k}, x) < \frac{\delta}{2} \quad \text{for all } k > N$$

and

$$d(x_{n_k}, y_{n_k}) < \varepsilon + \frac{1}{n_k} < \varepsilon + \frac{\delta}{2} \quad \text{for all } k > N.$$

Then

$$d(x, x_{n_k}) + d(x_{n_k}, y_{n_k}) < \frac{\delta}{2} + \varepsilon + \frac{\delta}{2} = \varepsilon + \delta = \rho(A, B) \leq d(x, y_{n_k}) \quad \text{for all } n_k > N.$$

But this contradicts the triangle inequality. Hence $\rho(x, B) = \rho(A, B)$.

Exercise 4.15. Let A be a compact subset of a metric space (X, d) and let B be a closed subset of X such that $A \cap B = \emptyset$. Show that $\rho(A, B) > 0$.

Proof. Since $\rho(A,B) = \inf\{d(x,y) : x \in A, y \in B\} \geq 0$, it is sufficient to show that $\rho(A,B) \neq 0$. Suppose contrary that $\rho(A,B) = 0$. Then by Exercise 4.14, there exists $x \in A$ such that $\rho(x,B) = \rho(A,B) = 0$. By Theorem 1.5, $B = \overline{B} = \{x \in X : \rho(x,B) = 0\}$. Therefore, $x \in B$ contradicting to our assumption that $A \cap B = \emptyset$.

This page intentionally left blank

Chapter 5
Continuous Functions

This chapter deals with the concepts of continuity and uniform continuity of functions defined on a metric space. Several characterizations and properties are presented. The concepts of homeomorphism and equivalent metrics are given. The uniform convergence of the sequences of functions defined on a metric space is discussed.

5.1 Definition and Characterizations

Definition 5.1. Let (X,d) and (Y,ρ) be metric spaces. A function $f : X \to Y$ is said to be *continuous at a point* $x_0 \in X$ if for every $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $x \in X$,

$$d(x,x_0) < \delta \quad \text{implies} \quad \rho(f(x),f(x_0)) < \varepsilon$$

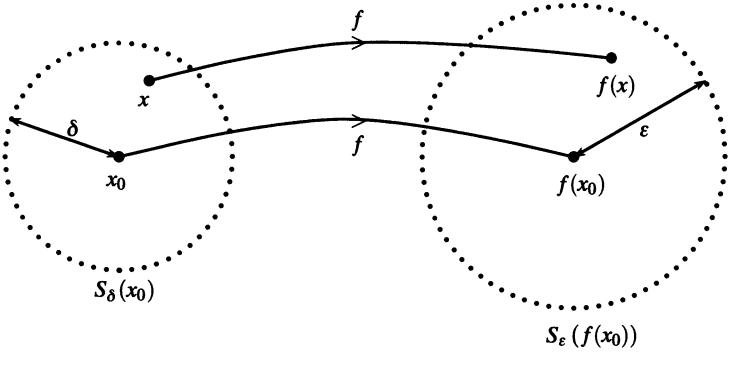
that is,

$$x \in S_\delta(x_0) \quad \text{implies} \quad f(x) \in S_\varepsilon(f(x_0)).$$

In other words, f is continuous at a point $x_0 \in X$ if for every $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$f(S_\delta(x_0)) \subseteq S_\varepsilon(f(x_0)).$$

The function f is said to be *continuous on* X if it is continuous at every point of X.



Example 5.1. (a) Let (X,d) be a metric space. The identity function $I : X \to X$ is continuous on X.

(b) Let $X = \mathbb{R}$ be the set of all real numbers with the usual metric. Then every constant function defined on X is continuous.

(c) Let X be a discrete metric space and Y be a metric space. Then every function $f : X \to Y$ is continuous on X.

(d) Let $X = Y = C[0,1]$ with the uniform metric $d_\infty(f,g) = \sup_{t\in[0,1]} |f(t) - g(t)|$. The mapping $\phi : X \to Y$ defined by

$$(\phi(f))(t) = \int_0^t f(s)ds \quad \text{for all } f \in C[0,1]$$

is continuous. In fact,

$$\begin{aligned} |(\phi(f))(t) - (\phi(g))(t)| &= \left| \int_0^t f(s)ds - \int_0^t g(s)ds \right| \\ &\le \int_0^t |f(s) - g(s)|\, ds \\ &\le d_\infty(f,g). \end{aligned}$$

Hence

$$d_\infty(\phi(f), \phi(g)) \le d_\infty(f,g) \quad \text{for all } f,g \in C[0,1].$$

So, the assertion follows if we choose $\delta = \varepsilon$.

Theorem 5.1. *Let (X,d) and (Y,ρ) be metric spaces. A function $f : X \to Y$ is continuous at a point $x_0 \in X$ if and only if for every sequence $\{x_n\} \subset X$, we have $x_n \to x_0$ implies $f(x_n) \to f(x_0)$.*

Proof. Let f be continuous at a point $x_0 \in X$. Then for a given $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $x \in X$,

$$d(x,x_0) < \delta \quad \text{implies} \quad \rho(f(x), f(x_0)) < \varepsilon.$$

Let $\{x_n\} \subset X$ be a sequence in X such that $x_n \to x_0$ as $n \to \infty$. Then there exists a positive integer N such that

$$d(x_n,x_0) < \delta \quad \text{for all } n > N.$$

Hence for all $n > N$, we have

$$\rho(f(x_n), f(x_0)) < \varepsilon,$$

and therefore, $f(x_n) \to f(x_0)$.

Conversely, assume that for every sequence $\{x_n\}$ in X such that $x_n \to x_0$, we have $f(x_n) \to f(x_0)$. Suppose that f is not continuous at x_0. Then there exists an $\varepsilon > 0$ such that for every $\delta > 0$, there is an $x \neq x_0$ satisfying

$$d(x,x_0) < \delta \quad \text{but} \quad \rho(f(x), f(x_0)) \ge \varepsilon.$$

In particular, for $\delta = \frac{1}{n}$ there is an x_n satisfying

$$d(x_n,x_0) < \frac{1}{n} \quad \text{but} \quad \rho(f(x_n), f(x_0)) \ge \varepsilon.$$

Then clearly $x_n \to x_0$ but $\{f(x_n)\}$ does not converge to $f(x_0)$. This contradicts to our hypothesis that $f(x_n) \to f(x_0)$. Hence f is continuous at x_0.

Theorem 5.2. *Let (X,d) and (Y,ρ) be metric spaces. A function $f:X\to Y$ is continuous on X if and only if for each $x\in X$ and for every sequence $\{x_n\}$ in X, we have $x_n\to x$ implies $f(x_n)\to f(x)$.*

Remark 5.1. If $f:X\to Y$ is continuous and $f(x_n)\to f(x)$ for any sequence $\{x_n\}$ in X, then the assertion $x_n\to x$ may not be true.

For example, let $X=\mathbb{R}$ be a metric space with the usual metric and $f:X\to X$ be a function defined as

$$f(x)=x^2 \quad \text{for all } x\in X.$$

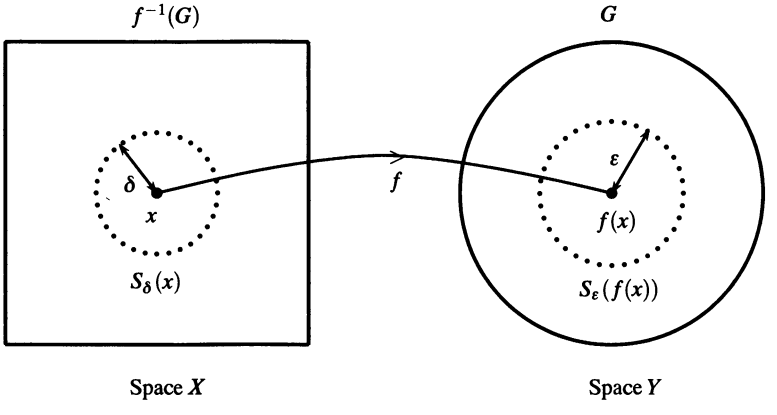
Consider a sequence $\{x_n\}$ in X, where $x_n=(-1)^n$ for all $n\in\mathbb{N}$. Then $f(x_n)=1\to f(1)$ as $n\to\infty$ but the sequence x_n does not converge to 1 as $n\to\infty$.

Theorem 5.3. *Let X and Y be metric spaces. A function $f:X\to Y$ is continuous on X if and only if $f^{-1}(G)$ is open in X whenever G is open in Y.*

Proof. Assume that f is continuous on X and G is an open set in Y. Then we prove that $f^{-1}(G)$ is open in X. If $f^{-1}(G)=\emptyset$, then the result is proved. So, we assume that $f^{-1}(G)\neq\emptyset$. Let $x\in f^{-1}(G)$, then $f(x)\in G$. Since G is open, $S_\varepsilon(f(x))\subseteq G$ for some $\varepsilon>0$. By the continuity of f, there exists a $\delta>0$ such that

$$f(S_\delta(x))\subseteq S_\varepsilon(f(x)),$$

and since $S_\varepsilon(f(x))\subseteq G$, it follows that $f(S_\delta(x))\subseteq G$. Thus $S_\delta(x)\subseteq f^{-1}(G)$ and hence $f^{-1}(G)$ is open.



Conversely, assume that $f^{-1}(G)$ is open in X whenever G is open in Y. Let $x\in X$ be arbitrary and $\varepsilon>0$ be given. Then $f(x)\in Y$ and $S_\varepsilon(f(x))$ is an open set. Therefore, by assumption $f^{-1}(S_\varepsilon(f(x)))$ is open in X and $x\in f^{-1}(S_\varepsilon(f(x)))$. Consequently, there exists a $\delta>0$ such that $S_\delta(x)\subseteq f^{-1}(S_\varepsilon(f(x)))$ and thus $f(S_\delta(x))\subseteq S_\varepsilon(f(x))$. Hence f is continuous at x. Since $x\in X$ was arbitrary, f is continuous on X.

Theorem 5.4. *Let X and Y be metric spaces. A function $f:X\to Y$ is continuous on X if and only if $f^{-1}(F)$ is closed in X whenever F is closed in Y.*

Proof. Let f be a continuous function and F be a closed in Y. Then $Y \setminus F$ is open in Y, and therefore, $f^{-1}(Y \setminus F)$ is open in X. Since

$$f^{-1}(F) = X \setminus \left(f^{-1}(Y \setminus F)\right)$$

and $f^{-1}(Y \setminus F)$ is open, it follows that $f^{-1}(F)$ is closed.

Conversely, assume that $f^{-1}(F)$ is closed in X whenever F is closed in Y. Then we show that f is continuous. Let G be an open subset of Y. Then $Y \setminus G$ is closed in Y and by hypothesis $f^{-1}(Y \setminus G)$ is closed in X. Since

$$f^{-1}(G) = X \setminus \left(f^{-1}(Y \setminus G)\right)$$

and $f^{-1}(Y \setminus G)$ is closed , we have $f^{-1}(G)$ is open. Hence f is continuous on X.

Remark 5.2. If f is a continuous function from a metric space X to another metric space Y, then the image $f(G)$ of an open set G in X need not be open in Y and the image $f(F)$ of a closed set F in X need not be closed in Y.

For example, consider the function $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = x^2$. Then, of course, f is continuous on $\mathbb{R}$. Let $G = (-1,1)$ be an open set in $\mathbb{R}$ but $f(G) = [0,1)$ is not open in $\mathbb{R}$.

Consider another function $f : [1,+\infty) \to \mathbb{R}$ defined by $f(x) = \frac{1}{x}$. Then f is continuous on $[1,+\infty)$. Let $A = [1,+\infty)$, then A is closed as a whole space with the usual metric but $f(A) = (0,1]$ is not closed in $\mathbb{R}$.

Theorem 5.5. *Let X and Y be metric spaces. A function $f : X \to Y$ is continuous if and only if $f\left(\overline{A}\right) \subseteq \overline{f(A)}$ for every subset A of X.*

Proof. Let f be a continuous function. Then $f^{-1}\left(\overline{f(A)}\right)$ is closed in X, since $\overline{f(A)}$ is closed in Y. Now we have

$$f(A) \subseteq \overline{f(A)} \;\Rightarrow\; A \subseteq f^{-1}\left(\overline{f(A)}\right) \;\Rightarrow\; \overline{A} \subseteq \overline{f^{-1}\left(\overline{f(A)}\right)}$$

and thus $\overline{A} \subseteq f^{-1}\left(\overline{f(A)}\right)$ because $f^{-1}\left(\overline{f(A)}\right)$ is closed. Hence $f\left(\overline{A}\right) \subseteq \overline{f(A)}$.

Conversely, let $f\left(\overline{A}\right) \subseteq \overline{f(A)}$ for every subset A of X. We prove that f is continuous. Let F be any closed set in Y. Then $\overline{F} = F$. Now we have

$$f\left(\overline{f^{-1}(F)}\right) \subseteq \overline{f\left(f^{-1}(F)\right)} = \overline{F} = F.$$

This implies that $\overline{f^{-1}(F)} \subseteq f^{-1}(F)$. But $f^{-1}(F) \subseteq \overline{f^{-1}(F)}$, therefore $\overline{f^{-1}(F)} = f^{-1}(F)$. Hence $f^{-1}(F)$ is closed in X. Thus f is continuous on X.

By combining Theorems 5.2 – 5.5, we have the following result.

Theorem 5.6. *Let X and Y be metric spaces and $f : X \to Y$ be a function. The following statements are equivalent:*

(a) f is continuous on X.
(b) For each $x \in X$ and for every sequence $\{x_n\}$ in X such that $x_n \to x$ implies $f(x_n) \to f(x)$.

(c) $f^{-1}(G)$ *is open in* X *wherever* G *is open in* Y.
(d) $f^{-1}(F)$ *is closed in* X *wherever* F *is closed in* Y.
(e) $f(\overline{A}) \subseteq \overline{f(A)}$ *for every subset* A *of* X.

Theorem 5.7. *Let* X, Y *and* Z *be metric spaces and* $f : X \to Y$ *and* $g : Y \to Z$ *be continuous functions. Then the composite function* $g \circ f : X \to Z$ *is continuous on* X.

Proof. Let G be an open set in Z. Then by continuity of g, $g^{-1}(G)$ is open in Y. The continuity of f implies that $f^{-1}\left(g^{-1}(G)\right)$ is open in X, that is, $\left(f^{-1} \circ g^{-1}\right)(G) = (g \circ f)^{-1}(G)$ is open in X. Hence $g \circ f$ is continuous.

Exercise 5.1. Let $f : X \to Y$ be a continuous function from a metric space X to another metric space Y. Give an example to show that $f(A^\circ) \subset (f(A))^\circ$ does not hold.
Hint: Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = \sin x$ for all $x \in \mathbb{R}$ and take $A = (0, 3\pi)$.

Exercise 5.2. Give an example to show that the following statement does not hold:

$f : X \to Y$ is continuous if and only if $(f(A))^\circ \subset f(A^\circ)$ for all subsets A of X.

Solution. The following function $f : \mathbb{R} \to \mathbb{R}$ defined as

$$f(x) = \begin{cases} 0 & \text{if } x \text{ is rational} \\ 1 & \text{if } x \text{ is irrational} \end{cases}$$

is not continuous. However, $f(A) \subset \{0,1\}$ for all $A \subset X$ and $(f(A))^\circ = \emptyset$. So, $(f(A))^\circ \subset f(A^\circ)$ holds for all subsets A of X.

For the converse, define a function $f : \mathbb{R} \to \mathbb{R}$ as $f(x) = |x|$ for all $x \in \mathbb{R}$ and consider the set

$$A = ([-2,-1] \cap \mathbb{Q}) \cup (-1,1) \cup ([1,2] \cap (\mathbb{R} \setminus \mathbb{Q})).$$

Then $A^\circ = (-1,1)$, and therefore, $f(A^\circ) = [0,1)$. But $f(A) = [0,2]$, and therefore, $(f(A))^\circ = (0,2)$.

Exercise 5.3. Let X and Y be metric spaces. Prove that the function $f : X \to Y$ is continuous on X if and only if for every subset B of Y, $\overline{f^{-1}(B)} \subseteq f^{-1}\left(\overline{B}\right)$.

Proof. Let f be continuous and $A = f^{-1}(B)$. Then by Theorem 5.5, we have $f\left(\overline{A}\right) \subseteq \overline{f(A)}$ and

$$f(A) \subseteq B \;\Rightarrow\; \overline{f(A)} \subseteq \overline{B} \;\Rightarrow\; f\left(\overline{A}\right) \subseteq \overline{B}.$$

Therefore, $\overline{A} \subseteq f^{-1}\left(\overline{B}\right)$ and thus $\overline{f^{-1}(B)} \subseteq f^{-1}\left(\overline{B}\right)$.

Conversely, let $\overline{f^{-1}(B)} \subseteq f^{-1}\left(\overline{B}\right)$ for every subset B of Y. Let F be a closed set in Y. Then $\overline{F} = F$ and by the hypothesis, we have

$$\overline{f^{-1}(F)} \subseteq f^{-1}\left(\overline{F}\right) = f^{-1}(F).$$

But $f^{-1}(F) \subseteq \overline{f^{-1}(F)}$, and therefore, $f^{-1}(F) = \overline{f^{-1}(F)}$. Thus $f^{-1}(F)$ is closed and hence f is continuous on X.

Exercise 5.4. Let X and Y be metric spaces and $f : X \to Y$ be a function. Prove that f is continuous on X if and only if for every subset B of Y, $f^{-1}(B^\circ) \subseteq \left[f^{-1}(B)\right]^\circ$.

Proof. Suppose that f is a continuous function on X. Let B be any arbitrary subset of Y. Then B° is open in Y and by continuity of f, $f^{-1}(B^\circ)$ is open in X. Therefore, $\left[f^{-1}(B^\circ)\right]^\circ = f^{-1}(B^\circ)$. But

$$B^\circ \subseteq B \Rightarrow f^{-1}(B^\circ) \subseteq f^{-1}(B)$$

and thus,

$$\left[f^{-1}(B^\circ)\right]^\circ \subseteq \left[f^{-1}(B)\right]^\circ.$$

This implies that $f^{-1}(B^\circ) \subseteq \left[f^{-1}(B)\right]^\circ$.

Conversely, let G be an open subset of Y. Then $G^\circ = G$. By the hypothesis

$$\left[f^{-1}(G)\right]^\circ \supseteq f^{-1}(G^\circ) = f^{-1}(G).$$

But $\left[f^{-1}(G)\right]^\circ \subseteq f^{-1}(G)$, therefore, $f^{-1}(G) = \left[f^{-1}(G)\right]^\circ$ and thus $f^{-1}(G)$ is open in X. Hence f is continuous on X.

Exercise 5.5. Let (X,d) be a metric space and for each $i = 1,2,\ldots,n$, the function $f_i : X \to \mathbb{R}$ be continuous. Then prove that the function $f : X \to \mathbb{R}^n$ defined by

$$f(x) = (f_1(x), f_2(x), \ldots, f_n(x)) \quad \text{for all } x \in X$$

is continuous.

Proof. Let x_0 be an arbitrary point in X. Since each f_i is continuous, for a given $\varepsilon > 0$, there exists $\delta_i > 0$ for each $i = 1,2,\ldots,n$ such that for all $x \in X$ and for each $i = 1,2,\ldots,n$

$$d(x,x_0) < \delta_i \quad \text{implies} \quad |f_i(x) - f_i(x_0)| < \frac{\varepsilon}{\sqrt{n}}.$$

Take $\delta = \min\{\delta_1, \delta_2, \ldots, \delta_n\}$. Then $d(x,x_0) < \delta$ implies that

$$d(f(x), f(x_0)) = \left\{(f_1(x) - f_1(x_0))^2 + \cdots + (f_n(x) - f_n(x_0))^2\right\}^{1/2} < \varepsilon.$$

Thus f is continuous at $x_0 \in X$. Since x_0 was arbitrary, f is continuous on X.

Exercise 5.6. Let (X,d) be a metric space and $f : X \to \mathbb{R}^n$ be a function defined as

$$f(x) = (f_1(x), f_2(x), \ldots, f_n(x)) \quad \text{for all } x \in X,$$

where $f_i : X \to \mathbb{R}$ is a function for each $i = 1,2,\ldots,n$. If f is continuous on X, then prove that each f_i is also continuous on X.

Hint: In fact, for each $i = 1,2,\ldots,n$

$$|f_i(x) - f_i(y)| \leq d(f(x), f(y)),$$

where $d(f(x), f(y)) = \left(\sum_{i=1}^{n} |f_i(x) - f_i(y)|^2\right)^{1/2}$.

5.2 Continuous Functions and Compact Spaces

Theorem 5.8. *Let X and Y be metric spaces and $f : X \to Y$ be a continuous function. If A is a compact subset of X, then $f(A)$ is compact in Y.*

Proof. Let $\mathscr{F} = \{G_\alpha\}_{\alpha \in \Lambda}$ be an open cover of $f(A)$. Then by Theorem 5.3, $f^{-1}(G_\alpha)$ is open in X for each $\alpha \in \Lambda$. Hence $\{A \cap f^{-1}(G_\alpha)\}_{\alpha \in \Lambda}$ forms an open cover of A. Since A is compact, there exists a finite set $J = \{1, 2, \ldots, n\}$ of Λ such that

$$A = \bigcup_{k=1}^{n} (A \cap f^{-1}(G_k)) = A \cap \left(\bigcup_{k=1}^{n} f^{-1}(G_k) \right) = A \cap f^{-1} \left(\bigcup_{k=1}^{n} G_k \right).$$

Thus

$$A \subseteq f^{-1} \left(\bigcup_{k=1}^{n} G_k \right)$$

and so

$$f(A) \subseteq \bigcup_{k=1}^{n} G_k.$$

Hence $\{G_1, G_2, \ldots, G_n\}$ is a finite subcover of $\mathscr{F}$. Thus $f(A)$ is compact.

Corollary 5.1. *Let X and Y be metric spaces and $f : X \to Y$ be a continuous function. If X is compact, then $f(X)$ is bounded.*

Proof. By Theorem 5.8, $f(X)$ is compact. Since every compact space is sequentially compact and every sequentially compact space is totally bounded, we have $f(X)$ is totally bounded and hence it is bounded.

Theorem 5.9. *Let X and Y be metric spaces and $f : X \to Y$ be a continuous function. If X is compact, then $f(F)$ is closed in Y whenever F is closed in X.*

Proof. Let F be a closed subset of X. Since every closed subset of a compact set is compact, by Theorem 5.8, we have $f(F)$ is compact and hence it is closed.

Theorem 5.10. *Let X and Y be metric spaces and $f : X \to Y$ be a continuous function. If f is bijective and X is compact, then f^{-1} is continuous on Y.*

Proof. Since f is bijective, $f^{-1} : Y \to X$ exists and also bijective. Let F be a closed set in X. Then $\left(f^{-1}\right)^{-1}(F) = f(F)$ and by Theorem 5.9, $f(F)$ is closed in Y. Thus, the inverse image of a closed set is closed and hence f^{-1} is continuous.

Remark 5.3. In Theorem 5.10, if X is not compact, then f^{-1} need not be continuous. For example, consider and identity function $I : (\mathbb{R}, d) \to (\mathbb{R}, \mathscr{U})$ from $\mathbb{R}$ with the discrete metric to $\mathbb{R}$ with the usual metric. Then I is continuous but I^{-1} is not.

Exercise 5.7. Give an example to show that Theorem 5.9 does not hold if X is not compact.

Exercise 5.8. Prove that a continuous real-valued function defined on a compact set is bounded and it assumes maximum and minimum values.

Proof. Let $f : X \to \mathbb{R}$ be continuous and let A be a compact subset of a metric space X. By Theorem 5.8, $f(A)$ is a compact subset of $\mathbb{R}$. By Heine-Borel Theorem 4.6, $f(A)$ is closed and bounded. Thus $\sup f(A)$ and $\inf f(A)$ exist and belong to $f(A)$. Therefore, there exist $\hat{x}, \tilde{x} \in A$ such that for all $y \in A$, $\inf f(A) = f(\hat{x}) \leq f(y) \leq f(\tilde{x}) = \sup f(A)$.

Exercise 5.9. Let (X,d) be a metric space and A be a nonempty compact subset of X. Prove that for every $x_0 \in X$, there exists a $y_0 \in A$ such that

$$d(x_0,y_0) = d(x_0,A) = \inf_{y \in A} d(x_0,y).$$

Proof. Consider the real valued function $f : A \to \mathbb{R}_+$ defined by $f(x) = d(x,x_0)$ for all $x \in A$. Now $|f(x) - f(y)| = |d(x,x_0) - d(y,x_0)| \leq d(x,y)$, so f is continuous on A. But A is compact, so f has a minimum on A by Exercise 5.8. That is, there exists a $y_0 \in A$ such that

$$f(y_0) = d(x_0,y_0) = \inf_{y \in A} d(x_0,y) = d(x_0,A).$$

5.3 Uniform Continuous Functions

Before giving the definition of uniform continuity, we examine the following examples.

Consider a real-valued function $f : [-1,1] \to \mathbb{R}$ defined as $f(x) = x^2$. Let x, x_0 be any points of $[-1,1]$. Then

$$\begin{aligned} d(f(x),f(x_0)) &= |f(x) - f(x_0)| = |x^2 - x_0^2| \\ &= |x - x_0| \cdot |x + x_0| < \varepsilon \end{aligned}$$

whenever $|x - x_0| < \frac{1}{2}\varepsilon = \delta$, where δ is independent on the choice of x and x_0.

Thus for any $\varepsilon > 0$, there exists a $\delta = \frac{1}{2}\varepsilon$ such that for any $x, x_0 \in [-1,1]$, we have

$$d(f(x),f(x_0)) < \varepsilon \quad \text{whenever} \quad d(x,x_0) < \delta.$$

Now, if we consider the same function $f(x) = x^2$ defined on $\mathbb{R}$, that is, $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) = x^2$. Then for every real numbers x, x_0, we have

$$d(f(x),f(x_0)) = |x^2 - x_0^2| = |x - x_0| \cdot |x + x_0| < \varepsilon$$

whenever $|x - x_0| < \frac{\varepsilon}{|x+x_0|} = \delta$, where δ depends on ε and x_0.

In this way, we see that δ may depend not only on ε but also on x_0. Uniform continuity is essentially continuity plus the added condition that for each ε we can find a δ which works uniformly over the entire space, in the sense that it does not depend on x_0.

Definition 5.2. Let (X,d) and (Y,ρ) be metric spaces. A function $f : X \to Y$ is said to be *uniformly continuous* if for each $\varepsilon > 0$, there exists a $\delta > 0$ (depends only on ε) such that for every $x, y \in X$,

$$d(x,y) < \delta \quad \text{implies} \quad \rho(f(x),f(y)) < \varepsilon.$$

Remark 5.4. Every uniform continuous function is continuous but converse need not be true in general. For example, in the first example mentioned above, δ is independent on the choice of x and x_0, and therefore, it is uniformity continuous. But in the latter example, δ depends on ε and x_0, and hence, it is only continuous but not uniformly continuous.

Example 5.2. (a) Let X be a discrete metric space and Y be any metric space. Then any function $f : X \to Y$ is uniformly continuous.

(b) Let $X = (0,1)$ be a metric space with the metric induced by the usual metric on $\mathbb{R}$ and $Y = \mathbb{R}$ with the usual metric. The function $f : X \to Y$ defined as $f(x) = \frac{1}{x}$ for all $x \in X$, is not uniformly continuous.

Let $\varepsilon = \frac{1}{2}$ and δ be any positive number. Choose $x = \frac{1}{n}$ and $y = \frac{1}{n+1}$, where n is a positive integer such that $n > \frac{1}{\delta}$. Then

$$|x-y| = \left|\frac{1}{n} - \frac{1}{n+1}\right| = \frac{1}{n(n+1)} < \frac{1}{n} < \delta,$$

but $|f(x) - f(y)| = |n-(n+1)| = 1 > \varepsilon$. Thus whatever $\delta > 0$ may be, there exist x and y such that $|x-y| < \delta$ but $|f(x) - f(y)| > \frac{1}{2}$.

(c) No polynomial function of degree greater than 1 is uniformly continuous on the usual metric space $\mathbb{R}$. Note that any polynomial function is continuous.

(d) The logarithmic function is not uniformly continuous on the usual metric space $X = (0,\infty)$.

Theorem 5.11. *Let (X,d) and (Y,ρ) be metric spaces and $f : X \to Y$ be a continuous function. If X is compact then f is uniformly continuous.*

Proof. Let $\varepsilon > 0$ and $x \in X$ be arbitrary. Consider the image $f(x)$ of x and the open sphere $S_\varepsilon(f(x))$. Since f is continuous, $f^{-1}(S_\varepsilon(f(x)))$ is an open set in X. Consider the family $\mathscr{F} = \{f^{-1}(S_\varepsilon(f(x))) : x \in X\}$ of these open sets in X. Then clearly $\mathscr{F}$ is an open cover of X. Since X is compact, it is sequentially compact, and therefore, by Lebesgue's covering lemma (Theorem 4.11) there exists a Lebesgue number $\delta > 0$ for $\mathscr{F}$. Thus every open sphere of diameter less than δ will be contained in at least one member of $\mathscr{F}$ and, consequently, we have

$$S_{\delta/2}(x) \subseteq f^{-1}(S_\varepsilon(f(x))) \quad \Rightarrow \quad f\left(S_{\delta/2}(x)\right) \subseteq S_\varepsilon(f(x)).$$

Hence for each $\varepsilon > 0$, there exists a $\tilde{\delta} > 0$ (independent of x) such that

$$d(x,y) < \frac{\delta}{2} = \tilde{\delta} \quad \Rightarrow \quad \rho(f(x),f(y)) < \varepsilon.$$

Hence f is uniformly continuous.

Theorem 5.12. *Composition of two uniformly continuous functions is a uniformly continuous function.*

Theorem 5.13. *Let (X,d) and (Y,ρ) be metric spaces and $f:X\to Y$ be an uniformly continuous function. If $\{x_n\}$ is a Cauchy sequence in X, then $\{f(x_n)\}$ is also a Cauchy sequence in Y.*

Proof. Since f is uniformly continuous, for each $\varepsilon>0$, there exists a $\delta>0$ (only depends on ε) such that for all $x_1,x_2\in X$,

$$d(x_1,x_2)<\delta \quad \text{implies} \quad \rho(f(x_1),f(x_2))<\varepsilon.$$

In particular, we have

$$d(x_n,x_m)<\delta \quad \text{implies} \quad \rho(f(x_n),f(x_m))<\varepsilon. \tag{5.1}$$

Since $\{x_n\}$ is a Cauchy sequence in X, for a given $\delta>0$, there exists a positive integer N such that

$$d(x_n,x_m)<\delta \quad \text{for all } n,m\geq N. \tag{5.2}$$

From (5.1) and (5.2), we obtain

$$\rho(f(x_n),f(x_m))<\varepsilon \quad \text{for all } n,m\geq N.$$

Hence $\{f(x_n)\}$ is a cauchy sequence in Y.

The following example shows that a continuous function may not map a Cauchy sequence into a Cauchy sequence.

Example 5.3. Let $X=(0,\infty)$ with the induced usual metric on $\mathbb{R}$ and $Y=\mathbb{R}$ with the usual metric. The function $f:X\to Y$ defined by $f(x)=\frac{1}{x}$ for all $x\in X$, is continuous on X. Clearly, $\{x_n : x_n=\frac{1}{n}\}_{n\in\mathbb{N}}$ is a Cauchy sequence in X. But $\{f(\frac{1}{n})\}_{n\in\mathbb{N}}=\{n\}_{n=1}^{\infty}$ is not a Cauchy sequence in Y. Indeed, the absolute difference of any two distinct points is at least as large as 1.

Exercise 5.10. Show that the function $f(x)=e^x$ defined on the usual metric space $\mathbb{R}$ is not uniformly continuous.
Hint: There is no $\delta>0$ that guarantees $d(x,y)=|e^x-e^y|<1$ for all $x,y\in\mathbb{R}$ with $|x-y|<\delta$. Specially, for all $x,y\in\mathbb{R}$ with $x<y$, we have $\dfrac{e^y-e^x}{y-x}>e^x$. Then for any $\gamma\in(0,1)$, pick $x=-\ln\gamma$ and $y=\gamma-\ln\gamma$, so that $e^y-e^x>(y-x)e^x=1$.

Exercise 5.11. Let (X,d) be a metric space and A be a subset of X. Prove that the function $f:X\to\mathbb{R}$ defined by

$$f(x)=d(x,A) \quad \text{for all } x\in X$$

is uniformly continuous.

Proof. By the triangular inequality

$$d(x,a)\leq d(x,y)+d(y,a) \quad \text{for all } a\in A, x\in X.$$

By taking infimum, we obtain

$$\inf_{a\in A} d(x,a)\leq d(x,y)+\inf_{a\in A} d(y,a).$$

Therefore,

$$d(x,A) \leq d(x,y) + d(y,A)$$

and so

$$d(x,A) - d(y,A) \leq d(x,y) \quad \text{for all } x,y \in X.$$

By interchanging x and y, we obtain

$$d(y,A) - d(x,A) \leq d(y,x) = d(x,y).$$

Thus

$$|d(x,A) - d(y,A)| \leq d(x,y).$$

Therefore, for a given $\varepsilon > 0$, choosing a δ such that $0 < \delta < \varepsilon$, we have

$$|f(x) - f(y)| = |d(x,A) - d(y,A)| \leq d(x,y) < \delta < \varepsilon,$$

that is,

$$|f(x) - f(y)| < \varepsilon \quad \text{whenever} \quad d(x,y) < \delta.$$

Hence f is uniformly continuous on X.

5.4 Homeomorphism and Equivalent Metrics

Definition 5.3. Let X and Y be two metric spaces. A function $f : X \to Y$ is said to be a *homeomorphism* if

(i) f is bijective, that is, f is one-to-one and onto
(ii) f and f^{-1} both are continuous.

Two metric spaces X and Y are said to be *homeomorphic* if there exists a homeomorphism from X onto Y.

Example 5.4. (a) Let $[0,1]$ and $[0,2]$ be metric spaces with the usual metric and $f : [0,1] \to [0,2]$ be a function defined as $f(x) = 2x$ for all $x \in [0,1]$. Then f is a homeomorphism and the spaces $[0,1]$ and $[0,2]$ are homeomorphic.

(b) Let $X = (0,\infty)$ and $Y = \mathbb{R}$ be metric spaces such that Y is equipped with the usual metric and X with the usual metric induced from $\mathbb{R}$. Let $f : X \to Y$ be a function defined as $f(x) = \ln x$ for all $x \in X$. Then f is a homeomorphism between X and Y.

(c) The metric space $\mathbb{R}$ with the usual metric is not homeomorphic to the metric space $\mathbb{R}$ with the discrete metric.

Recall that a mapping $f : X \to Y$ from a metric space (X,d) to a metric space (Y,ρ) is an *isometry* if

$$\rho(f(x), f(y)) = d(x,y) \quad \text{for all } x,y \in X.$$

It is obvious that an isometry is one-to-one and uniformly continuous.

Recall also that X and Y are said to be *isometric* if there exists an isometry between them that is onto.

An isometry is necessarily a homeomorphism, but the converse need not be true. In Example 5.4 (a) and (b), the mappings are homeomorphism but not isometry.

Theorem 5.14. *Let X and Y be two metric spaces and $f : X \to Y$ be a bijective function. Then the following statements are equivalent.*

(a) f is a homeomorphism.
(b) The set $G \subset X$ is open if and only if its image $f(G) \subset Y$ is open.
(c) The set $F \subset X$ is closed if and only if its image $f(F) \subset Y$ is closed.

Proof. (a) $\Rightarrow$ (b). Let (a) hold. Then f and f^{-1} both are continuous. Therefor, for any open set $G \subset X$, we have $\left(f^{-1}\right)^{-1}(G) = f(G)$ is open in Y by using continuity of f^{-1}.

Conversely, let $f(G) \subset Y$ be open. Then by continuity of f, $f^{-1}(f(G)) = G$ is open in X ($f^{-1}(f(G)) = G$ because f is injective).

(b) $\Rightarrow$ (c). Let $F \subset X$ be a closed set. Then $X \setminus F$ is open in X. By (b), $f(X \setminus F)$ is open in Y. But $f(F) = Y \setminus f(X \setminus F)$. Hence $f(F)$ is closed in Y.

Conversely, let $f(F)$ is closed in Y. Then $Y \setminus f(F)$ is open in Y. Then $f(X \setminus F) = Y \setminus f(F)$ is open. By (b), $X \setminus F$ is open and hence F is closed.

(c) $\Rightarrow$ (a). Let F be a closed subset of X. Then by hypothesis, $f(F)$ is closed in Y. Since $f^{-1}(f(F)) = F$, the inverse image of closed set is closed, and therefore, f is continuous.

Also, since $\left(f^{-1}\right)^{-1}(F) = f(F)$, by hypothesis F is closed in X and $f(F)$ is closed in Y, we have f^{-1} is continuous on Y. Therefore, f is a homeomorphism.

Remark 5.5. Let X, Y and Z be metric spaces. If X is homeomorphic to Y and Y is homeomorphic to Z, then X is homeomorphic to Z.

Definition 5.4. Two metrics d_1 and d_2 on the same underlying set X are said to be *equivalent* if for every sequence $\{x_n\}$ in X and $x \in X$,

$$\lim_{n\to\infty} d_1(x_n, x) = 0 \quad \text{if and only if} \quad \lim_{n\to\infty} d_2(x_n, x) = 0,$$

that is, a sequence converges to x with respect to the metric d_1 if and only if it converges to x with respect to the metric d_2.

The metric spaces (X, d_1) and (X, d_2) are said to be *equivalent* if the metrics d_1 and d_2 are equivalent.

Remark 5.6. (i) Two metrics d_1 and d_2 on a nonempty set X are equivalent if and only if the identity map $I : (X, d_1) \to (X, d_2)$ together with its inverse map $I^{-1} : (X, d_2) \to (X, d_1)$ are continuous, that is, if and only if the identity map $I : (X, d_1) \to (X, d_2)$ is a homeomorphism.

(ii) If two metrics are equivalent then the families of open sets are same in (X, d_1) and (X, d_2).

The following result provides a sufficient condition for two metrics on a set to be equivalent.

Theorem 5.15. *Two metrics d_1 and d_2 on a nonempty set X are equivalent if there exists a constant k such that*

$$\frac{1}{k} d_2(x, y) \le d_1(x, y) \le k d_2(x, y) \quad \text{for all } x, y \in X. \tag{5.3}$$

Proof. Let $\{x_n\}$ be a sequence in X such that $d_1(x_n,x) \to 0$ as $n \to \infty$. From the first inequality in (5.3), we have

$$\frac{1}{k}d_2(x_n,x) \leq d_1(x_n,x) \to 0 \quad \text{as } n \to \infty$$

and thus, $d_2(x_n,x) \to 0$ as $n \to \infty$.

If $d_2(x_n,x) \to 0$ as $n \to \infty$, then in view of second inequality in (5.3), we have

$$d_1(x_n,x) \leq kd_2(x_n,x) \to 0 \quad \text{as } n \to \infty$$

and thus, $d_1(x_n,x) \to 0$ as $n \to \infty$.

The following example shows that the converse of the above theorem does not hold, that is, if d_1 and d_2 are equivalent metrics then the inequality (5.3) need not be true.

Example 5.5. Let (X,d) be a metric space and ρ be anther metric on X defined by

$$\rho(x,y) = \frac{d(x,y)}{1+d(x,y)} \quad \text{for all } x,y \in X.$$

Then d and ρ are equivalent metrics on X.

Indeed, let $d(x_n,x) \to 0$ as $n \to \infty$. Then obviously $\rho(x_n,x) \to 0$ as $n \to \infty$. On the other hand, if $\rho(x_n,x) \to 0$ as $n \to \infty$, then for $0 < \varepsilon < 1$, we have $2-\varepsilon > 1$, and also, there exists a positive integer N such that $\rho(x_n,x) < \varepsilon/2$ for all $n > N$. So

$$\frac{d(x_n,x)}{1+d(x_n,x)} = \rho(x_n,x) < \frac{\varepsilon}{2} \quad \text{for all } n > N$$

which is equivalent to

$$d(x_n,x) < \frac{\varepsilon}{2-\varepsilon} < \varepsilon \quad \text{for all } n > N.$$

Thus, $d(x_n,x) \to 0$ as $n \to \infty$.

On the other hand, we observe that there is no constant $k > 0$ such that

$$\frac{1}{k}d(x,y) \leq \rho(x,y) \quad \text{for all } x,y \in \mathbb{R}.$$

This establishes that the condition (5.3) in Theorem 5.15 is not necessary for the metrics d and ρ to be equivalent.

This example shows that there always exists an equivalent metric ρ on a metric space such that (X,ρ) is bounded.

Now we present some examples of equivalent and non-equivalent metrics.

Example 5.6. (a) The metrics d_1, d_2 and d_∞ defined on $\mathbb{R}^n$ by

$$d_1(x,y) = \sum_{i=1}^{n} |x_i - y_i|$$

$$d_2(x,y) = \left(\sum_{i=1}^{n}(x_i - y_i)^2\right)^{1/2}$$

$$d_\infty(x,y) = \max_{1\le i\le n}|x_i - y_i|$$

for all $x = (x_1,\ldots,x_n) \in \mathbb{R}^n$ and $y = (y_1,\ldots,y_n) \in \mathbb{R}^n$, are equivalent. In fact, any real numbers $\alpha_1, \alpha_2, \ldots, \alpha_n$, we have

$$\left(\sum_{i=1}^{n}\alpha_i^2\right)^{1/2} \le \sum_{i=1}^{n}|\alpha_i| \le n\cdot\max\{|\alpha_i| : 1\le i\le n\} \le n\left(\sum_{i=1}^{n}\alpha_i^2\right)^{1/2}.$$

(b) Let $C[0,1]$ be a space of all continuous linear functionals defined on $[0,1]$. Consider the metrics d_∞ and d defined on $C[0,1]$ by

$$d_\infty(f,g) = \sup_{x\in[0,1]}|f(x) - g(x)|$$

and

$$d(f,g) = \int_0^1 |f(x) - g(x)|\,dx$$

for all $f, g \in C[0,1]$. Then these metrics d_∞ and d are not equivalent.

Indeed, let $\{f_n\}$ be a sequence in $C[0,1]$ defined by

$$f_n(x) = x^n \quad \text{for all } x \in [0,1].$$

Also, let $f(x) = 0$ for all $x \in [0,1]$. Then $d(f_n, f) \to 0$ as $n \to \infty$. In fact,

$$d(f_n,f) = \int_0^1 x^n dx = \frac{1}{n+1} \to 0 \quad \text{as } n \to \infty.$$

However,

$$d_\infty(f_n,f) = \sup_{x\in[0,1]}|f_n(x) - f(x)| = 1 \quad \text{for all } n.$$

So, $\lim_{n\to\infty} d_\infty(f_n,f) \neq 0$. Thus the sequence $\{f_n\}$ converges to f with respect to the metric d but not with respect to the metric d_∞. Therefore, the metrics d_∞ and d are not equivalent.

However, for any sequence $\{f_n\}$ in $C[0,1]$ and any $f \in C[0,1]$, when $d_\infty(f_n,f) \to 0$ as $n \to \infty$, it does follow that $d(f_n,f) \to 0$ as $n \to \infty$. Observe that

$$d(f_n,f) = \int_0^1 |f_n(x) - f(x)|\,dx \le \sup_{x\in[0,1]}|f_n(x) - f(x)| = d_\infty(f_n,f).$$

(c) Let m be the space of all bounded sequences with the following two metrics

$$d(x,y) = \sup_{1\le n<\infty}|x_n - y_n|$$

and

$$d_1(x,y) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{|x_n - y_n|}{1 + |x_n - y_n|},$$

where the sequences $x = (x_1, x_2, \ldots)$ and $y = (y_1, y_2, \ldots)$ are elements of m.

Consider the sequence $\{e_n\}$ in m whose respective terms are the sequences

$$e_1 = (1,0,0,\ldots), e_2 = (0,1,0,\ldots), \ldots, e_n = (0,0,\ldots,0,1,0,\ldots), \ldots$$

where 1 is in the nth place. Let $e_0 = (0,0,0,\ldots)$. Then

$$d(e_n, e_0) = 1 \quad \text{while} \quad d_1(e_n, e_0) = \frac{1}{2^n} \frac{1}{1+1} = \frac{1}{2^{n+1}} \to 0 \text{ as } n \to \infty.$$

Thus the sequence $\{e_n\}$ converges with respect to the metric d_1 but not with respect to the metric d. Hence the metrics d and d_1 are not equivalent.

Exercise 5.12. Let X, Y and Z be metric spaces. If X is homeomorphic to Y and Y is homeomorphic to Z, then prove that X is homeomorphic to Z.

Proof. Let $f : X \to Y$ and $g : Y \to Z$ be homeomorphisms. Then f, g are bijective, continuous and f^{-1}, g^1 are also continuous. Consider the function $g \circ f : X \to Z$.

Since composition of two bijective and continuous functions is bijective and continuous, respectively, $g \circ f$ is bijective and continuous. Also since $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$ is continuous. Therefore, $g \circ f$ is a homeomorphism and hence X is homeomorphic to Z.

Exercise 5.13. Show that the function $f : \mathbb{R} \to (-1,1)$ defined by

$$f(x) = \frac{x}{1+|x|} \quad \text{for all } x \in \mathbb{R}$$

is a homeomorphism. Also, show that f is uniformly continuous on $\mathbb{R}$.

Hint: The given function f can be restated as

$$f(x) = \begin{cases} \frac{x}{(1+x)} & \text{if } x \geq 0, \\ \frac{x}{(1-x)} & \text{if } x < 0. \end{cases}$$

Observe that the inverse function $g : (-1,1) \to \mathbb{R}$ of the function f is given by

$$g(y) = \begin{cases} \frac{y}{(1-y)} & \text{if } y \geq 0, \\ \frac{y}{(1+y)} & \text{if } y < 0. \end{cases}$$

In fact, for $x \geq 0$, we have $\frac{x}{(1+x)} \geq 0$ and hence,

$$(g \circ f)(x) = g\left(\frac{x}{1+x}\right) = \frac{\frac{x}{(1+x)}}{1 - \frac{x}{(1+x)}} = x,$$

whereas, for $x < 0$, we have $\frac{x}{(1-x)} < 0$ and hence,

$$(g \circ f)(x) = g\left(\frac{x}{1-x}\right) = \frac{\frac{x}{(1-x)}}{1 - \frac{x}{(1-x)}} = x.$$

Similar computations for $f \circ g$ is left for the readers. Since the function f and its inverse g are both continuous, f is a homeomorphism.

Now we check the uniform continuity of f. For $x, y > 0$,

$$|f(x) - f(y)| = \left| \frac{x}{1+x} - \frac{y}{1+y} \right| = \frac{|x-y|}{(1+x)(1+y)} < |x-y|.$$

Similarly, for $x < 0$ and $y < 0$, we have $|f(x) - f(y)| < |x-y|$. For $x > 0$ and $y < 0$,

$$\begin{aligned} |f(x) - f(y)| &= \left| \frac{x}{1+x} - \frac{y}{1+y} \right| = \frac{|x-y-2xy|}{(1+x)(1-y)} \\ &\leq \frac{x-y-2xy+x^2+y^2}{(1+x)(1-y)} \\ &= \frac{(x-y)(1+x-y)}{(1+x)(1-y)} \\ &\leq x-y = |x-y|. \end{aligned}$$

The case when x or y is 0 is left to the readers.

Exercise 5.14. (a) Show that the image of a complete metric space under homeomorphism need not be complete.
(b) Show that the image of a complete metric space under a one-to-one uniformly continuous mapping need not be complete.

Hint: The function f in Exercise 5.13 is both homeomorphism and uniformly continuous. $\mathbb{R}$ is complete but $f(\mathbb{R}) = (-1, 1)$ is not.

Exercise 5.15. A mapping $f : X \to Y$ from a metric space (X, d) to a metric space (Y, ρ) is said to be *open* (respectively, *closed*) if $f(G)$ is open (respectively, closed) in Y for every open (respectively, closed) set G in X.

(a) Given an example of a function f that is continuous and open, but is not a homeomorphism.
(b) Given an example of a function f that is continuous and closed, but is not a homeomorphism.
(c) Given an example of a function f that is continuous and closed, but is not open.
(d) Given an example of a function f that is continuous and open, but is not closed.

Hint: (a) Consider $S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ and define a map $f : \mathbb{R} \to S$ as $f(t) = (\cos(2\pi t), \sin(2\pi t))$. Since the components of f are continuous, and therefore, f is continuous. Observe that it is also open. Indeed, let $(a, b) = I$ and $\ell(I) = b - a$. If $\ell(I) > 1$, then $f(I) = S$ which is open in S. If $\ell(I) < 1$, then $f(I)$ is an arc without endpoints and so is open in S, being the intersection of S with an open disc with the missing endpoints on its edge. The function is not one-to-one and thus, cannot be a homeomorphism.

(b) and (c) Consider any constant function $f : \mathbb{R}^2 \to \mathbb{R}$ and observe that it is continuous and closed, but not a homeomorphism, as it is not one-to-one.

(d) Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x_1, x_2) = x_1$ for all $(x_1, x_2) \in \mathbb{R}^2$ and observe that it is continuous and open but not closed, as it maps the hyperbola $x_1 x_2 = 1$ onto the set $\mathbb{R} \setminus \{0\}$, which is not closed.

5.5 Uniform Convergence of Sequences of Functions

Definition 5.5. Let $\{f_n\}$ be a sequence of functions from a metric space (X,d) to another metric space (Y,ρ) and $f : X \to Y$ be a function. We say that the sequence $\{f_n\}$ *converges pointwise* to the function f if for a given $\varepsilon > 0$ and given $x \in X$, there exists a positive integer N such that

$$\rho(f_n(x), f(x)) < \varepsilon \quad \text{for all } n > N. \tag{5.4}$$

In general, the integer N depends on ε as well as on x. It is not always possible to find an integer N such that (5.4) holds for all $x \in X$ simultaneously.

If N depends only on ε, then we say that the sequence $\{f_n\}$ converges uniformly on X.

Definition 5.6. Let $\{f_n\}$ be a sequence of functions from a metric space (X,d) to another metric space (Y,ρ) and $f : X \to Y$ be a function. We say that the sequence $\{f_n\}$ *converges uniformly* to the function f if for a given $\varepsilon > 0$, there exists a positive integer N (depending only on ε) such that

$$\rho(f_n(x), f(x)) < \varepsilon \quad \text{for all } n > N \text{ and all } x \in X. \tag{5.5}$$

It is clear that uniform convergence implies pointwise convergence, but converse need not be true.

Example 5.7. Let $X = [0,1]$ and $Y = \mathbb{R}$ with the usual metric. Let $\{f_n\}$ be a sequence of functions $f_n : X \to Y$ defined by

$$f_n(x) = x^n.$$

Then $\{f_n\}$ converges pointwise to the function $f : X \to Y$ defined by

$$f(x) = \begin{cases} 0 & \text{if } 0 \le x < 1 \\ 1 & \text{if } x = 1. \end{cases}$$

Remark 5.7. (i) Observe that the function f is not continuous even though each of the function f_n is continuous.

(ii) Since f is not continuous and $\{f_n\}$ is a sequence of continuous functions, by next Theorem 5.16, $\{f_n\}$ does not converge unifromly to f.

Example 5.8. Let $X = Y = \mathbb{R}$ with the usual metric. Let $\{f_n\}$ be a sequence of functions $f_n : X \to Y$ defined by

$$f_n(x) = \begin{cases} 1 - \frac{1}{n}|x| & \text{if } |x| < n \\ 0 & \text{if } |x| \ge n. \end{cases}$$

Then $\{f_n\}$ converges pointwise to the function $f : X \to Y$ defined by

$$f(x) = 1.$$

But $\{f_n\}$ does not converge uniformly to f.

Indeed, let $\varepsilon = \frac{1}{2}$. Note that for every $n \in \mathbb{N}$, there exists $x_0 \in X$ with $f_n(x_0) = 0$ and so

$$|f_n(x_0) - f(x_0)| = 1 > \varepsilon.$$

Theorem 5.16. *Let (X,d) and (Y,ρ) be metric spaces, $\{f_n\}$ be a sequence of continuous functions each defined on X to Y and $f : X \to Y$ be a function. If $f_n \to f$ uniformly on X, then f is continuous on X. In other words, an uniform limit of a sequence of continuous functions is continuous.*

Proof. Let $x_0 \in X$ be arbitrary and let $\varepsilon > 0$ be given. Since $f_n \to f$ uniformly on X, there exists a positive integer N (depending on ε only) such that for each $x \in X$,

$$\rho(f_n(x), f(x)) < \frac{\varepsilon}{3} \quad \text{for all } n > N. \tag{5.6}$$

Let $n_0 > N$. Since f_{n_0} is continuous at x_0, we can choose $\delta > 0$ such that for all $x \in S_\delta(x_0)$, that is, $d(x,x_0) < \delta$, we have

$$\rho\left(f_{n_0}(x), f_{n_0}(x_0)\right) < \frac{\varepsilon}{3}. \tag{5.7}$$

Using (5.6) and (5.7), for all $x \in S_\delta(x_0)$, we have

$$\begin{aligned}\rho(f(x), f(x_0)) &\le \rho\left(f(x), f_{n_0}(x)\right) + \rho\left(f_{n_0}(x), f_{n_0}(x_0)\right) + \rho\left(f_{n_0}(x_0), f(x_0)\right) \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon,\end{aligned}$$

and therefore, f is continuous at x_0.

Theorem 5.17 (Cauchy Criterion). *Let $\{f_n\}$ be a sequence of functions defined on a metric space (X,d) to a complete metric space (Y,ρ). Then there exists a continuous function $f : X \to Y$ such that*

$$f_n \to f \text{ uniformly on } X$$

if and only if the following condition is satisfied: For every $\varepsilon > 0$, there exists a positive integer N such that

$$\rho(f_m(x), f_n(x)) < \varepsilon \quad \text{for all } m,n > N \text{ and for all } x \in X.$$

Proof. Suppose that $f_n \to f$ uniformly on X. Then for any given $\varepsilon > 0$, there exists a positive integer N such that

$$\rho(f_n(x), f(x)) < \frac{\varepsilon}{2} \quad \text{for all } n > N \text{ and for all } x \in X.$$

Thus for all $m,n > N$ and for all $x \in X$,

$$\rho(f_m(x), f(x)) < \frac{\varepsilon}{2} \quad \text{and} \quad \rho(f_n(x), f(x)) < \frac{\varepsilon}{2}.$$

Hence for all $m,n > N$ and for all $x \in X$, we have

$$\rho(f_m(x), f_n(x)) < \rho(f_m(x), f(x)) + \rho(f_n(x), f(x)) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Conversely, assume that

$$\rho(f_m(x), f_n(x)) < \varepsilon \quad \text{for all } m,n > N \text{ and for all } x \in X.$$

Then for each $x \in X$, the sequence $\{f_n(x)\}$ is a Cauchy sequence in a complete metric space Y, and therefore, converges to some point $f(x)$, say. We claim that $f_n \to f$ uniformly on X. If $\varepsilon > 0$ is given, we can choose N such that for all $n > N$, we have

$$\rho(f_n(x), f_{n+k}(x)) < \frac{\varepsilon}{2} \quad \text{for all } k = 1,2,\ldots \text{ and all } x \in X.$$

Letting $k \to \infty$, we obtain

$$\rho(f_n(x), f(x)) = \lim_{n\to\infty} \rho(f_n(x), f_{n+k}(x)) \leq \frac{\varepsilon}{2}.$$

Hence

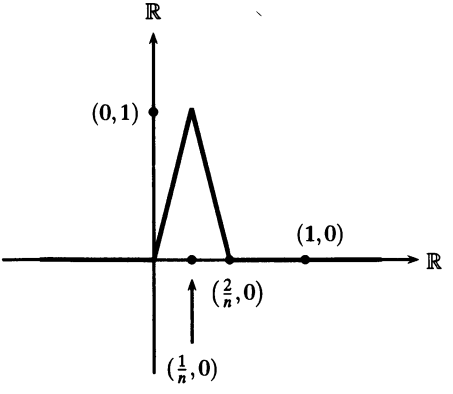
$$\rho(f_n(x), f(x)) < \varepsilon \quad \text{for all } n > N \text{ and for all } x \in X,$$

and so $f_n \to f$ uniformly on X.

Exercise 5.16. Let $X = Y = \mathbb{R}$ with the usual metric. Let $\{f_n\}$ be a sequence of functions defined by

$$f_n(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ nx & \text{if } 0 \leq x \leq \frac{1}{n} \\ -nx+2 & \text{if } \frac{1}{n} \leq x \leq \frac{2}{n} \\ 0 & \text{if } x \geq \frac{2}{n} \end{cases}$$

and $f(x) = 0$ for all $x \in \mathbb{R}$. Show that $f_n \to f$ pointwise on $\mathbb{R}$.



Proof. If x is close to 0 and positive, then there exists a positive integer N such that $2/N < x$ and so $f_n(x) = 0$ for all $n > N$. However, there exists no N such that (5.5) holds for $\varepsilon = \frac{1}{2}$, for all $n > N$ and for all $x \in \mathbb{R}$. For, if such an N were to exist, we would have

$$f_n(x) < \frac{1}{2} \quad \text{for all } n > N \text{ and for all } x \in \mathbb{R}.$$

We would then have $Nx < \frac{1}{2}$ for $0 \le x \le \frac{1}{N}$; but for $x = \frac{2}{3}N$, we obtain the contradiction $\frac{2}{3} < \frac{1}{2}$.

Exercise 5.17. Prove that the sequence $\{f_n\}$ defined by

$$f_n(x) = \tan^{-1}(nx) \quad \text{for all } x \ge 0,$$

is uniformly convergent on $[\alpha, \infty)$ when $\alpha > 0$, but is not uniformly convergent on $[0, \infty)$.

Proof. The pointwise limit function is

$$f(x) = \lim_{n\to\infty} f_n(x) = \begin{cases} \frac{\pi}{2} & \text{if } x > 0 \\ 0 & \text{if } x = 0. \end{cases}$$

We show that $f_n \to f$ uniformly on $[\alpha, \infty)$ when $\alpha > 0$.

For $x > 0$, we claim that

$$|f_n(x) - f(x)| = \left|\tan^{-1}(nx) - \frac{\pi}{2}\right| = \cot^{-1}(nx).$$

Since $0 < \tan^{-1}\theta < \pi/2$ for any $\theta > 0$, therefore, when $x > 0$, we have $0 < \tan^{-1}(nx) < \pi/2$ and hence

$$0 < \frac{\pi}{2} - \tan^{-1}(nx) < \frac{\pi}{2}. \tag{5.8}$$

Also,

$$\cot\left(\frac{\pi}{2} - \tan^{-1}(nx)\right) = nx. \tag{5.9}$$

It follows from (5.8) and (5.9) that $\pi/2 - \tan^{-1}(nx) = \cot^{-1}(nx)$. The first inequality in (5.8) implies that

$$\left|\tan^{-1}(nx) - \frac{\pi}{2}\right| = \frac{\pi}{2} - \tan^{-1}(nx) \quad \text{for all } x > 0.$$

Thus

$$\left|\tan^{-1}(nx) - \frac{\pi}{2}\right| = \cot^{-1}(nx).$$

Let $\varepsilon > 0$ be arbitrary. When $x \ge \alpha$, the inequality $n > \frac{\cot\varepsilon}{\alpha}$ implies that $n > \frac{\cot\varepsilon}{x}$, so that $nx > \cot\varepsilon$ and hence $\cot^{-1} nx < \varepsilon$ in view of the fact that $\cot^{-1}$ is a decreasing function. It follows that if N is an integer greater than or equal to $\frac{\cot\varepsilon}{\alpha}$, then

$$|f_n(x) - f(x)| = \left|\tan^{-1}(nx) - \frac{\pi}{2}\right| = \cot^{-1} nx < \varepsilon$$

whenever $n > N$ and $x \ge \alpha$. However, $\frac{\cot\varepsilon}{x} \to \infty$ as $n \to 0$, so that no integer N exists for which

$$|f_n(x) - f(x)| < \varepsilon \quad \text{for all } n > N \text{ and all } x \in [0, \infty).$$

Actually this proves that the convergence fails to be uniform even on the smaller set $(0,\infty)$.

Exercise 5.18. Consider the sequence $\{f_n\}$ of functions defined by $f_n(x) = \dfrac{1}{1+x^n}$ for $x \in [0,1]$.

(a) Find $f(x) = \lim\limits_{n\to\infty} f_n(x)$.
(b) Show that for $0 < a < 1$, $\{f_n\}$ converges uniformly to f on $[0,a]$.
(c) Show that $\{f_n\}$ does not converge uniformly to f on $[0,1]$.

Proof. (a) Since $f_n(1) = 1/2$ for all $n \in \mathbb{N}$, we obtain $f(1) = 1/2$ and

$$\lim_{n\to\infty} \frac{1}{1+x^n} = 1 \quad \text{for } 0 \le x < 1$$

which yields $f(x) = 1$ for $0 \le x < 1$.

(b) If $x \in [0,a]$, then

$$\left| \frac{1}{1+x^n} - 1 \right| = \frac{x^n}{1+x^n} \le \frac{a^n}{1+a^n}.$$

Since $0 < a < 1$,

$$\lim_{n\to\infty} \frac{a^n}{1+a^n} = 0,$$

so $\{f_n\}$ converges uniformly to f on $[0,a]$.

(c) Given $n \in \mathbb{N}$, let x be such that $\sqrt[n]{\dfrac{1}{2}} < x < 1$, then $\dfrac{1}{2} < x^n < 1$. Therefore,

$$\left| \frac{1}{1+x^n} - 1 \right| = \frac{x^n}{1+x^n} > \frac{\frac{1}{2}}{1+1} = \frac{1}{4}$$

which implies $\{f_n\}$ does not converge uniformly to f on $[0,1]$.

Exercise 5.19. Let $\{f_n\}$ be a sequence of functions defined by $f_n(x) = \dfrac{nx}{e^{nx}}$ for $x \in [0,2]$.

(a) Show that $\lim\limits_{n\to\infty} f_n(x) = 0$ for $x \in (0,2]$.
(b) Show that the convergence is not uniform on $[0,2]$.

Proof. (a) Since

$$\frac{f_{n+1}(x)}{f_n(x)} = \frac{(n+1)xe^{nx}}{nxe^{(n+1)x}} = \frac{(n+1)}{n} e^{-x},$$

we have $\lim\limits_{n\to\infty} \dfrac{f_{n+1}(x)}{f_n(x)} = e^{-x} < 1$. Thus $\lim\limits_{n\to\infty} f_n(x) = 0$ for $x \in (0,2]$.

(b) Let us find the maximum of f_n in $[0,2]$. Since $f_n'(x) = \dfrac{ne^{nx} - ne^{nx}nx}{(e^{nx})^2} = 0$, we have

$$e^{nx}(n - n^2x) = 0 \;\Leftrightarrow\; n^2x = n \;\Leftrightarrow\; x = \frac{1}{n}.$$

It can be easily seen that $\frac{1}{n}$ is the maximum with $f_n\left(\frac{1}{n}\right) = \frac{1}{e}$. Since

$$\lim_{n\to\infty} \sup\{|f_n(x) - 0| : x \in [0,2]\} = \lim_{n\to\infty} f_n\left(\frac{1}{n}\right) = \frac{1}{e},$$

the convergence is not uniform on $[0,2]$.

Exercise 5.20. Let $X = [0,1]$ and $Y = \mathbb{R}$ with the usual metric. Let $\{f_n\}$ be a sequence of functions defined by

$$f_n(x) = \begin{cases} 4n^2x & \text{if } 0 \le x \le \frac{1}{2n} \\ -4n^2x + 4n & \text{if } \frac{1}{2n} < x < \frac{1}{n} \\ 0 & \text{if } \frac{1}{n} \le x \le 1. \end{cases}$$

Show that the sequence $\{f_n\}$ converges pointwise to the constant function $f(x) = 0$.

Chapter 6
Connected Spaces

In this chapter, we discuss the connectedness, components and their properties. The idea of connected spaces or sets comes from an interval which is in one piece. That is, an interval cannot be written as a disjoint union of two nonempty open (in it) subintervals.

6.1 Separated Sets

Definition 6.1. Let X be a metric space. Two nonempty subsets A and B of X are said to be *separated* if

$$A \cap \overline{B} = \emptyset \text{ and } \overline{A} \cap B = \emptyset.$$

Example 6.1. 1. In the usual metric space $\mathbb{R}$,

(a) the sets $A = (1,2)$ and $B = (2,3)$ are separated;
(b) the sets $A = [1,2]$ and $B = (2,3)$ are not separated.

2. In the discrete metric space $\mathbb{R}$, the sets $A = (1,2]$ and $B = (2,3)$ are separated, since in a discrete metric space X, $\overline{A} = A$.

Remark 6.1. If A and B are separated sets in a metric space X, then they are disjoint, since $A \cap B \subseteq A \cap \overline{B} = \emptyset$. But two disjoint sets need not be separated. For example, $A = (1,2]$ and $B = (2,3)$ are disjoint but not separated in the usual metric space $\mathbb{R}$. Also, the sets $A = \{x \in \mathbb{R} : x < 0\}$ and $B = \{x \in \mathbb{R} : x \geq 0\}$ are disjoint but not separated in the usual metric space $\mathbb{R}$.

Theorem 6.1. *Let (X,d) be a metric space and, A and B be nonempty subsets of X. If $d(A,B) > 0$, then A and B are separated.*

Proof. Let $d(A,B) = \lambda$. Since

$$d(A,B) = \inf\{d(x,y) : x \in A, y \in B\} = \lambda > 0,$$

we have

$$d(x,y) \geq \lambda \quad \text{for all } x \in A \text{ and } y \in B.$$

If $x \in A$, then x cannot be a limit point of B because $S_{\frac{\lambda}{2}}(x)$ contains no points of B. Therefore, $A \cap \overline{B} = \emptyset$. Similarly it can be proved that $\overline{A} \cap B = \emptyset$. Hence A and B are separated.

Remark 6.2. The converse of the above theorem is not true in general.

Example 6.2. In the usual metric space $\mathbb{R}$, the sets

$$A = \{x \in \mathbb{R} : x > 0\} \text{ and } B = \{x \in \mathbb{R} : x < 0\}$$

are separated, while $d(A,B) = 0$.

Theorem 6.2. *Let A and B be two nonempty subsets of a metric space X.*

(a) If A and B are closed, then they are separated if and only if A and B are disjoint.
(b) If A and B are open, then they are separated if and only if A and B are disjoint.

Proof. Since any two separated sets are disjoint, we only prove one way.
(a) Since A and B are closed, we have

$$\overline{A} \cap B = A \cap \overline{B} = A \cap B.$$

By the definition of separated sets, we follow the result.

(b) Let A and B be open and disjoint, that is, $A \cap B = \emptyset$. Then $X \setminus A$ and $X \setminus B$ are closed and thus

$$X \setminus A = \overline{X \setminus A}, \quad X \setminus B = \overline{X \setminus B}$$

and

$$A \subseteq X \setminus B \quad \text{implies} \quad \overline{A} \subseteq \overline{X \setminus B} = X \setminus B.$$

Similarly, we can show that $\overline{B} \subseteq X \setminus A$. Therefore,

$$\overline{A} \cap B \subseteq (X \setminus B) \cap B = \emptyset \quad \text{and} \quad \overline{B} \cap A \subseteq (X \setminus A) \cap A = \emptyset.$$

Hence A and B are separated.

Remark 6.3. Theorem 6.2 says that, in a metric space, any two sets, either both open or both closed are separated if and if only if they are disjoint.

Theorem 6.3. *Let A and B be two separated sets is a metric space X and $G = A \cup B$. Then*

(a) A and B are open if G is open;
(b) A and B are closed if G is closed.

Proof. (a) Since $\overline{B}$ is closed, $(\overline{B})^c$ is open. Since $G = A \cup B$ is open, so is $(A \cup B) \cap (\overline{B})^c$. Since A and B are separated, $A \cap \overline{B} = \emptyset$ and hence $A \subseteq (\overline{B})^c$. Also, $B \subseteq \overline{B}$ implies $(\overline{B})^c \subseteq B^c$ and thus $B \cap (\overline{B})^c \subseteq B \cap B^c = \emptyset$. Consequently,

$$(A \cup B) \cap (\overline{B})^c = [A \cap (\overline{B})^c] \cup [B \cap (\overline{B})^c] = A \cup \emptyset = A.$$

Hence A is open. Similarly, we can prove that B is open.

(b) Since $G = A \cup B$ is closed, we have $A \cup B = \overline{A \cup B} = \overline{A} \cup \overline{B}$. By hypothesis, A and B are separated and thus $\overline{A} \cap B = \emptyset$. Hence

$$\overline{A} = \overline{A} \cap (\overline{A} \cup \overline{B}) = \overline{A} \cap (A \cup B) = (\overline{A} \cap A) \cup (\overline{A} \cap B) = A \cup \emptyset = A.$$

Thus A is closed. Similarly, we can prove that B is closed.

Exercise 6.1. Let C and D be two separated subsets of a metric space X and let $A \subseteq C$ and $B \subseteq D$. Prove that A and B are also separated.

6.2 Connected Sets

Definition 6.2. A metric space X is said to be *connected* if it cannot be expressed as the union of two separated sets.

Furthermore, X is called *disconnected* if it is not connected, that is, if X can be expressed as the union of two separated sets. In other words, X is disconnected if

$$X = A \cup B \quad \text{where} \quad A \cap \overline{B} = \emptyset \text{ and } \overline{A} \cap B = \emptyset.$$

The sets A and B are said to form a *separation* of X.

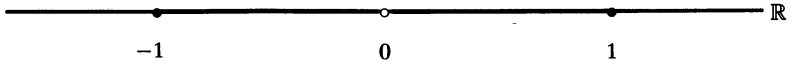
Remark 6.4. In view of Theorem 6.2, we have

(i) a metric space X is connected if it cannot be expressed as the union of two nonempty disjoint closed sets;
(ii) a metric space X is connected if it cannot be expressed as the union of two nonempty disjoint open sets.

Definition 6.3. A subset Y of a metric space (X,d) is said to be *connected* if the subspace (Y,d_Y) with the metric induced from X is connected.

In other words (Y,d_Y) is connected if $Y \neq A \cup B$, where A and B are nonempty disjoint subsets of Y which are open in Y, that is, $A = G \cap Y$ and $B = H \cap Y$, where G and H are open subsets of X.

Example 6.3. Let $Y = [-1,0) \cup (0,1]$ be the subspace of the usual metric space $\mathbb{R}$. Each of the sets $[-1,0)$ and $(0,1]$ is nonempty and open in Y (although not open in $\mathbb{R}$). Therefore, they form a separation of Y. Alternatively, note that neither of these sets contains a limit point of the other. (They do have a limit point 0 in common, but that does not matter). Thus Y is disconnected. It can also be seen in the following figure.



Example 6.4. (a) Let $\mathbb{Q}$ be the set of all rational numbers with the usual metric $d(x,y) = |x-y|$ for all $x,y \in \mathbb{Q}$. Then Q is disconnected, because the sets

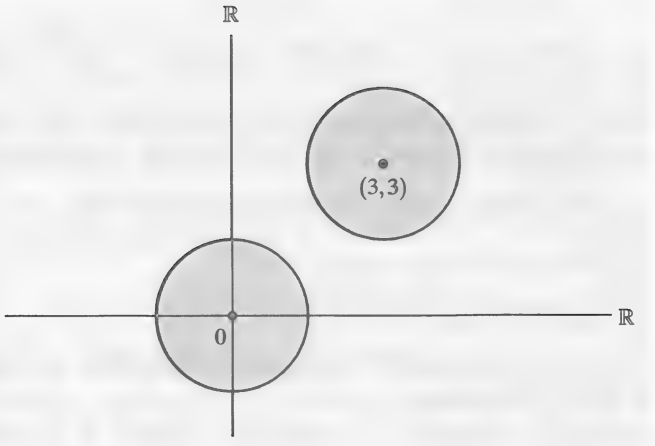
$$A = \left\{x \in \mathbb{Q} : x > \sqrt{2}\right\} \quad \text{and} \quad B = \left\{x \in Q : x < \sqrt{2}\right\}$$

are open in $\mathbb{Q}$ and nonempty disjoint sets such that $A \cap \overline{B} = \emptyset, \overline{A} \cap B = \emptyset$ and $\mathbb{Q} = A \cup B$.

Similarly, the set of all irrational numbers is disconnected.

(b) The space $X = \left\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\right\} \cup \left\{(x,y) \in \mathbb{R}^2 : (x-3)^2 + (y-3)^2 \leq 1\right\}$ with the usual metric induced from $\mathbb{R}^2$ is disconnected.

(c) A singleton set $\{x\}$ in any metric space X is always connected.



Theorem 6.4. *A subset A of the usual metric space $\mathbb{R}$ is connected if and only if it is an interval.*

Proof. Let A be a connected subset of $\mathbb{R}$. Suppose contrary that A is not an interval. Then there exist real numbers x,y and z such that $x < z < y$ and $x,y \in A$ but $z \notin A$. Then A can be written as

$$A = [A \cap (-\infty,z)] \cup [A \cap (z,+\infty)]$$

and

$$[A \cap (-\infty,z)] \cap [A \cap (z,+\infty)] = \emptyset.$$

Since $x \in A \cap (-\infty,z)$ and $y \in A \cap (z,+\infty)$, $A \cap (-\infty,z)$ and $A \cap (z,+\infty)$ are nonempty; also they are clearly disjoint and open in A. Thus A is the union of two disjoint nonempty open sets, and therefore, A is disconnected which is a contradiction of our assumption. Hence A is an interval.

Conversely, assume that A is an interval but is not connected. Then there exist two nonempty sets G and H closed in A such that

$$A = G \cup H \quad \text{and} \quad G \cap H = \emptyset.$$

Since G and H are nonempty, we can choose $x \in G$ and $y \in H$. Also since G and H are disjoint, we have $x \neq y$. We may assume that $x < y$. Since A is an interval, $[x,y] \subseteq A$ and each point in $[x,y]$ belongs to either G or H. We now define z by

$$z = \sup([x,y] \cap G) = \sup\{t \in G : x \leq t \leq y\}.$$

Then clearly $z \in [x,y]$ and so $z \in A$. Since G is closed and by the definition of z, we have $z \in G$. This implies that $z < y$. Again, by the definition of z, $z + \varepsilon \in H$ for each $\varepsilon > 0$ such that $z + \varepsilon \leq y$. Since H is closed in A, we have $z \in H$. Thus z belongs to both G and H which contradicts to our assumption that these sets are disjoint. Hence A is connected.

Corollary 6.1. *The real line $\mathbb{R}$ is connected.*

Proof. Since $\mathbb{R}$ is an interval $(-\infty, \infty)$, the assertion of the corollary follows from Theorem 6.4.

Theorem 6.5. *A metric space X is connected if and only if the only nonempty subset of X which is both open and closed is X itself.*

Proof. Let X be connected. Assume that A is a nonempty proper subset of X which is both open and closed. Let $B = X \setminus A$. Then clearly $X = A \cup B$ and A and B are nonempty, disjoint and open subsets of X. Hence X is disconnected which contradicts to our hypothesis. Thus X is the only nonempty set which is both open and closed.

Conversely, let X be the only nonempty set which is both open and closed. Suppose that X is disconnected. Then $X = A \cup B$, where A and B are nonempty, disjoint and open subsets of X. This implies that $B = X \setminus A$. Since A is open, B is closed as a complement of the open set A. Therefore, B is open as well as closed proper subset of X which contradicts to our hypothesis. Hence X is connected.

Theorem 6.6. *Let Y be a connected subset of a metric space X such that $Y \subseteq A \cup B$, where A and B are separated sets in X. Then either $Y \subseteq A$ or $Y \subseteq B$, that is, Y cannot intersect both A and B.*

Proof. Since $Y \subseteq A \cup B$, we have $Y \cap (A \cup B) = Y$, and therefore, $(Y \cap A) \cup (Y \cap B) = Y$.

If $Y \cap A = \emptyset$, then $Y = \emptyset \cup (Y \cap B) = Y \cap B \subseteq B$, and therefore, $Y \subseteq B$.

If $Y \cap B = \emptyset$, then $Y = \emptyset \cup (Y \cap A) = Y \cap A \subseteq A$, and therefore, $Y \subseteq A$.

So it is sufficient to show that either $Y \cap A = \emptyset$ or $Y \cap B = \emptyset$. Suppose contrary that neither $Y \cap A = \emptyset$ nor $Y \cap B = \emptyset$. Since $Y \cap A$ and $Y \cap B$ are contained in the separated sets A and B, respectively, we obtain $Y \cap A$ and $Y \cap B$ are separated. Since $Y = (Y \cap A) \cup (Y \cap B)$, we have that Y is the union of two nonempty separated sets, and therefore, Y is disconnected which contradicts to our hypothesis. Hence either $Y \cap A = \emptyset$ or $Y \cap B = \emptyset$.

Theorem 6.7. *If M is a connected subset of a metric space X, then every subset N such that $M \subseteq N \subseteq \overline{M}$ is connected. In particular, $\overline{M}$ is connected.*

Proof. Suppose that N is not connected. Then there exist two nonempty sets A and B open in N such that

$$N = A \cup B \text{ and } A \cap B = \emptyset.$$

Since $M \subseteq N$ and, A and B are open in N, we have

$$A_0 = A \cap M \text{ and } B_0 = B \cap M$$

are nonempty and open in the metric subspace M. Moreover,

$$M = N \cap M = (A \cup B) \cap M = (A \cap M) \cup (B \cap M) = A_0 \cup B_0$$

and

$$A_0 \cap B_0 = (A \cap M) \cap (B \cap M) = (A \cap B) \cap M = \emptyset \cap M = \emptyset.$$

This contradicts the hypothesis that M is connected. Hence N is connected.

Alternative Proof of Theorem 6.7. Suppose that N is not connected. Then there exist two nonempty separated sets A and B such that $N = A \cup B$. Since $M \subseteq N$, we have $M \subseteq A \cup B$. Since the connected set M is contained in the union of two separated sets A and B, it follows from Theorem 6.6 that either $M \subseteq A$ or $M \subset B$. Let $M \subseteq A$ then $\overline{M} \subseteq \overline{A}$ and hence $\overline{M} \cap B \subseteq \overline{A} \cap B = \emptyset$ because A and B are separated. Therefore, $\overline{M} \cap B = \emptyset$.

Since $N = A \cup B$ and $N \subset \overline{M}$, we have $B \subseteq N \subseteq \overline{M}$ and hence $\overline{M} \cap B = B$. Since $\overline{M} \cap B = \emptyset$ and $\overline{M} \cap B = B$, we have $B = \emptyset$ which contradicts to the fact that B is nonempty. Hence N is connected.

Since $M \subseteq N \subseteq \overline{M}$, we have $M \subseteq \overline{M} \subseteq \overline{M}$. Taking $\overline{M} = N$, we have $\overline{M}$ is connected.

Theorem 6.8. *Let $\{A_\alpha\}_{\alpha \in \Lambda}$ be a family of connected sets in a metric space X such that $\bigcap_{\alpha \in \Lambda} A_\alpha \neq \emptyset$. Then*

$$A = \bigcup_{\alpha \in \Lambda} A_\alpha \text{ is connected.}$$

Proof. Suppose that A is not connected. Then there exist nonempty sets $P \subseteq A$ and $Q \subseteq A$ open in A such that

$$A = P \cup Q \quad \text{and} \quad P \cap Q = \emptyset.$$

Let $x \in \bigcap_{\alpha \in \Lambda} A_\alpha \subseteq A = P \cup Q$. Then either $x \in P$ or $x \in Q$. Assume that $x \in P$. Then $P \cap A_\alpha \neq \emptyset$ for each $\alpha \in \Lambda$. Moreover, there exists an index α_0 such that $P \cap A_{\alpha_0} \neq \emptyset$. Similarly, $Q \cap A_{\alpha_0} \neq \emptyset$. Consequently, the sets $P_0 = P \cap A_{\alpha_0}$ and $Q_0 = Q \cap A_{\alpha_0}$ are nonempty sets open in A_{α_0} such that

$$\begin{aligned} A_{\alpha_0} &= A \cap A_{\alpha_0} = (P \cup Q) \cap A_{\alpha_0} \\ &= (P \cap A_{\alpha_0}) \cup (Q \cap A_{\alpha_0}) \\ &= P_0 \cup Q_0 \end{aligned}$$

and

$$\begin{aligned} P_0 \cap Q_0 &= (P \cap A_{\alpha_0}) \cap (Q \cap A_{\alpha_0}) \\ &= (P \cap Q) \cap A_{\alpha_0} = \emptyset \cap A_{\alpha_0} = \emptyset. \end{aligned}$$

This contradicts the hypothesis that A_{α_0} is connected. Hence $A = \bigcup_{\alpha \in \Lambda} A_\alpha$ is connected.

Exercise 6.2. Let $\{A_\alpha\}_{\alpha \in \Lambda}$ be a family of connected subsets of a metric space X such that one of the member of this family intersects every other member. Then $A = \bigcup_{\alpha \in \Lambda} A_\alpha$ is connected.

Proof. Let A_{α_0} be a fixed member of the given family which intersects every other member, that is,

$$A_{\alpha_0} \cap A_\alpha \neq \emptyset \quad \text{for each } \alpha \in \Lambda.$$

Let $D_\alpha = A_{\alpha_0} \cup A_\alpha$ for each $\alpha \in \Lambda$. Since D_α is the union of two connected sets having nonempty intersection, it follows from Theorem 6.8 that D_α is connected for each $\alpha \in \Lambda$. Now

$$\bigcup_{\alpha\in\Lambda} D_\alpha = \bigcup_{\alpha\in\Lambda} (A_{\alpha_0}\cup A_\alpha) = A_{\alpha_0}\bigcup\left(\bigcup_{\alpha\in\Lambda} A_\alpha\right) = \bigcup_{\alpha\in\Lambda} A_\alpha,$$

because $A_{\alpha_0}\subseteq\bigcup_{\alpha\in\Lambda}A_\alpha$. For $\alpha\neq\beta$,

$$\begin{aligned} D_\alpha\cap D_\beta &= (A_{\alpha_0}\cup A_\alpha)\cap(A_{\alpha_0}\cup A_\beta)\\ &= A_{\alpha_0}\cup(A_\alpha\cap A_\beta). \end{aligned}$$

Since $A_{\alpha_0}\cap A_\alpha\neq\emptyset$ for each $\alpha\in\Lambda$, we have that $A_{\alpha_0}\neq\emptyset$ and hence $D_\alpha\cap D_\beta\neq\emptyset$ for each $\alpha\neq\beta$. This implies that $\{D_\alpha\}_{\alpha\in\Lambda}$ is the family of connected subsets of X having a nonempty intersection. By Theorem 6.8, $\bigcup_{\alpha\in\Lambda} D_\alpha$ is connected. Since $\bigcup_{\alpha\in\Lambda} D_\alpha = \bigcup_{\alpha\in\Lambda} A_\alpha$, $\bigcup_{\alpha\in\Lambda} A_\alpha$ is connected.

Exercise 6.3. Let X be a metric space and $\{A_n\}$ be a sequence of connected sets such that $A_n\cap A_{n+1}\neq\emptyset$ for all $n\in\mathbb{N}$. Prove that $A=\bigcup_{n\in\mathbb{N}} A_n$ is connected.

Proof. For each $n\in\mathbb{N}$, let $D_n=\bigcup_{i=1}^n A_i$. We prove the result by induction. If $n=1$, then $D_1=A_1$ is connected. Suppose that D_n is connected. Since $A_n\cap A_{n+1}\neq\emptyset$ and $A_n\cap A_{n+1}\subseteq D_n\cap A_{n+1}$, $D_n\cap A_{n+1}\neq\emptyset$. Therefore by Theorem 6.8, $D_{n+1}=A_{n+1}\cup D_n$ is connected. Since $\bigcap_{n\in\mathbb{N}} D_n = A_1$, $A=\bigcup_{n\in\mathbb{N}} A_n$ is connected.

Exercise 6.4. Let M be a subset of a metric space X. If every pair of points in M lies in a connected subset of M, then prove that M is connected.

Proof. Suppose contrary that M is not connected. Then there exist nonempty separated sets A and B such that $M=A\cup B$. Let $x\in A$ and $y\in B$, then $x,y\in M$. By hypothesis, there exists a connected set $M_{xy}\subset M$ such that $x,y\in M_{xy}\subset M$. Then either $M_{xy}\subset A$ or $M_{xy}\subset B$. Suppose that $M_{xy}\subset A$, then $y\in M_{xy}\subset A$. On the other hand, $y\notin A$ because A and B are separated. This is a contradiction. Hence M is connected.

Exercise 6.5. Show that the set $\mathbb{N}$ of all natural numbers with the metric induced by the usual metric on $\mathbb{R}$ is not connected.

Exercise 6.6. Prove that the subset $A=\{x\in\mathbb{R}:|x|>0\}$ of $\mathbb{R}$ is disconnected.

Proof. Since $A=\{x\in\mathbb{R}:|x|>0\}=(-\infty,0)\cup(0,\infty)$, A is the union of two nonempty disjoint open sets in A, that is,

$$A=[A\cap(-\infty,0)]\cup[A\cap(0,\infty)]$$

where $A\cap(-\infty,0)$ and $A\cap(0,\infty)$ are open in A, nonempty and disjoint. Hence A is disconnected.

Exercise 6.7. Show that the subset $A=\{(x,y)\in\mathbb{R}^2:x^2-y^2\geq 4\}$ of $\mathbb{R}^2$ is disconnected.
Hint: If $G=\{(x,y):x<-1\}$ and $H=\{(x,y):x>1\}$, then $G\cap A$ and $H\cap A$ are nonempty disjoint open sets in A whose union is A.

Exercise 6.8. Show that the closed or open sphere in $\mathbb{R}^n$ is connected.

Exercise 6.9. Let $X = \{(x,y) \in \mathbb{R}^2 : y = 0\} \cup \{(x,y) \in \mathbb{R}^2 : x > 0 \text{ and } y = \frac{1}{x}\}$ be a subset of the usual metric space $\mathbb{R}^2$. Prove that X is not connected.
Hint: The two indicated sets $\{(x,y) \in \mathbb{R}^2 : y = 0\}$ and $\{(x,y) \in \mathbb{R}^2 : x > 0 \text{ and } y = \frac{1}{x}\}$ form a separation of X, because neither contains a limit point of the other (prove it).

Exercise 6.10. Show that the connected subspaces of the set of rational numbers $\mathbb{Q}$ are the only one point sets.

Proof. Let Y be a connected subspace of $\mathbb{Q}$ containing two points p and q. Then we can choose an irrational number α lying between p and q. Then

$$Y = [Y \cap (-\infty, \alpha)] \cup [Y \cap (\alpha, +\infty)]$$

and

$$[Y \cap (-\infty, \alpha)] \cap [Y \cap (\alpha, +\infty)] = \emptyset,$$

and also $Y \cap (-\infty, \alpha)$ and $Y \cap (\alpha, +\infty)$ are open in Y and nonempty because $p \in Y \cap (-\infty, \alpha)$ and $q \in Y \cap (\alpha, +\infty)$. Hence Y is disconnected which contradicts to our hypothesis that Y is connected. Therefore, Y can not have more than one point.

Exercise 6.11. Let A and B be two connected subsets of a metric space X such that $\overline{A} \cap B \neq \emptyset$. Show that $A \cup B$ is connected.

6.3 Continuous Functions and Connected Sets

Let $X_0 = \{0,1\}$ and let d_0 denote the discrete metric on X_0. The metric space (X_0, d_0) is called a *discrete two element space.*

Theorem 6.9. *A metric space (X,d) is disconnected if and only if there exists a continuous mapping from (X,d) onto the discrete two element space (X_0, d_0).*

Proof. Let (X,d) be disconnected. Then there exist disjoint open sets A and B of X such that

$$X = A \cup B \quad \text{and} \quad A \cap B = \emptyset.$$

Define a mapping $f : X \to X_0$ by

$$f(x) = \begin{cases} 0 & \text{if } x \in A \\ 1 & \text{if } x \in B. \end{cases}$$

Clearly, $f^{-1}(\{0\}) = A$ and $f^{-1}(\{1\}) = B$. Therefore, f is onto. It remains to show that f is continuous from (X,d) to (X_0, d_0). The open subsets of the discrete two element space (X_0, d_0) are precisely, $\emptyset$, $\{0\}$, $\{1\}$ and $\{0,1\}$. Observe that $f^{-1}(\emptyset) = \emptyset$ and $f^{-1}(\{0,1\}) = X$ are open in (X,d). Moreover, $f^{-1}(\{0\}) = A$ and $f^{-1}(\{1\}) = B$ are open sets in (X,d). Hence f is continuous.

Conversely, let $f : (X,d) \to (X_0, d_0)$ be continuous and onto. Let $A = f^{-1}(\{0\})$ and $B = f^{-1}(\{1\})$. Then A and B are nonempty disjoint open in X such that $X = A \cup B$. Therefore, X is disconnected.

Corollary 6.2. *A metric space (X,d) is connected if and only if the only continuous mappings from (X,d) onto the discrete two element space (X_0,d_0) are the constant mappings, namely, the mappings $f(x)=1$ for all $x\in X$ or $g(x)=0$ for all $x\in X$.*

The continuous image of a connected space is connected. More precisely, we have the following result.

Theorem 6.10. *Let (X,d) be a connected metric space, (Y,ρ) a metric space and $f:(X,d)\to(Y,\rho)$ a continuous mapping. Then the space $f(X)$ with the metric induced from ρ is connected.*

Proof. By our hypothesis, the map $f:X\to f(X)$ is continuous. Assume contrary that $f(X)$ is not connected. Then by above theorem, there exists a continuous function $g:f(X)\to X_0$ from $f(X)$ onto the discrete two element space (X_0,d_0). The composite mapping $g\circ f:X\to X_0$ of two continuous mappings is also continuous from X onto X_0. Again, by above theorem, X is disconnected which is a contradiction of our hypothesis.

Corollary 6.3. *Let (X,d) be a connected metric space and $f:X\to\mathbb{R}$ be a continuous function. Then $f(X)$ is an interval.*

Proof. By Theorem 6.10, $f(X)$ is connected subset of $\mathbb{R}$ and the result follows from Theorem 6.4.

As an application of Theorem 6.4, we derive the following intermediate value theorem.

Theorem 6.11 (Intermediate Value Theorem). *If $f:[a,b]\to\mathbb{R}$ is a continuous function, then for every y such that $f(a)\le y\le f(b)$ or $f(b)\le y\le f(a)$, there exists $x\in[a,b]$ for which $f(x)=y$.*

Proof. It is sufficient to consider only the case when $f(a)\neq y\neq f(b)$. Let y be any real number such that $f(a)<y<f(b)$. By Theorem 6.4, $[a,b]$ is a connected subset of $\mathbb{R}$. Hence $f([a,b])$ is an interval by Theorem 6.4 and Corollary 6.3. Therefore, there exists an $x\in[a,b]$ such that $f(x)=y$. Similarly, we can prove the result when $f(b)<y<f(a)$.

The following result is the converse of the intermediate value theorem.

Theorem 6.12. *Let (X,d) be a metric space. If every continuous function $f:X\to\mathbb{R}$ has the intermediate value property (that is, if $y_1,y_2\in f(X)$ and y is a real number between y_1 and y_2, then there exists an $x\in X$ such that $f(x)=y$), then (X,d) is connected.*

Proof. Assume to the contrary that (X,d) is not connected. Then by Theorem 6.9, there exists a continuous mapping $g:(X,d)\to(X_0,d_0)$ that is onto. Define a mapping $h:(X_0,d_0)\to\mathbb{R}$ by $h(0)=0$ and $h(1)=1$. Then h is continuous. Consider the composite map $h\circ g:(X,d)\to\mathbb{R}$. Then clearly, $h\circ g$ is continuous as it is composition of two continuous functions. Also, $\{0,1\}\subseteq(h\circ g)(X)$. However, there exists no $x\in X$ such that $(h\circ g)(x)=1/2$. In fact, $(h\circ g)^{-1}(\{1/2\})=(g^{-1}\circ h^{-1})(\{1/2\})=g^{-1}(\emptyset)=\emptyset$.

The following fixed point theorem is an interesting application of the intermediate value theorem.

Theorem 6.13. *If $I = [-1,1]$ and $f : I \to I$ is continuous, then there exists a point $c \in I$ such that $f(c) = c$.*

Proof. If $f(-1) = -1$ or $f(1) = 1$, then we get the conclusion. So, we assume that $f(-1) > -1$ and $f(1) < 1$. Define

$$g(x) = f(x) - x \quad \text{for all } x \in I.$$

Then g is the difference of two continuous functions and so it is also continuous. The mapping g also satisfies the inequalities $g(-1) = f(-1) + 1 > 0$ and $g(1) = f(1) - 1 < 0$. By Theorem 6.11, there exists $c \in (-1,1)$ such that $g(c) = 0$, that is, $f(c) = c$.

Remark 6.5. The interval $[-1,1]$ can not be replaced by $[-1,1)$ or $[-1,\infty)$. In the former case, the function $f : [-1,1) \to [-1,1)$ defined by $f(t) = \frac{t+1}{2}$ for all $t \in [-1,1)$, is continuous but has no fixed point. In the latter case, we define $f : [-1,\infty) \to [-1,\infty)$ as $f(t) = t + 1$ for all $t \in [-1,\infty)$, then f is continuous but has no fixed point.

Exercise 6.12. Let X and Y be metric spaces and let $f : X \to Y$ be a continuous function. If C is connected subset of X, then $f(C)$ is connected subset of Y.

Proof. Assume that $f(C)$ is disconnected. Then $f(C) = G \cup H$, where G and H are nonempty, disjoint open subsets of Y such that $f(C) \cap G$ and $f(C) \cap H$ are nonempty. Then

$$C \subseteq f^{-1}(G \cup H) = f^{-1}(G) \cup f^{-1}(H).$$

Since f is continuous, $f^{-1}(G)$ and $f^{-1}(H)$ are open in X. Moreover, $C \cap f^{-1}(G)$ and $C \cap f^{-1}(G)$ are nonempty and disjoint. It follows that C is disconnected which is a contradiction. Hence $f(C)$ is connected.

6.4 Components

Definition 6.4. A maximal connected subset C of a metric space X is called a *component* of X. In other words, C is a component of a metric space X if C is connected and it is not properly contained in any larger connected subset of X.

Let X be a metric space and $x \in X$. The union $C(x)$ of all connected subsets containing x is clearly a maximal connected subset of X containing x. Such union $C(x)$ is called *connected component of* x in X.

Example 6.5. (a) The components of the subspace $Y = [-1,0) \cup (0,1]$ of the real line $\mathbb{R}$ are the two sets $[-1,0)$ and $(0,1]$.

(b) Let $\mathbb{Q}$ be the set of all rational numbers with the usual metric. The component of each $x \in \mathbb{Q}$ is the set consisting x alone. In other words, any subset A of $\mathbb{Q}$ containing more than one point is disconnected. Indeed, if $x, y \in A$, $x < y$, then $(-\infty, \alpha) \cap A$ and $(\alpha, \infty) \cap A$ provide a disconnection of A, when $x < \alpha < y$ and α is irrational.

Theorem 6.14. *Each component of a metric space X is closed.*

Proof. Suppose that the component C is not closed. Then C is properly contained in $\overline{C}$. By Theorem 6.7, $\overline{C}$ is connected. Therefore, C is properly contained in a connected subset of X which contradicts the maximality of C. Hence C is closed.

Remark 6.6. The components need not be open sets. For example, let

$$X = \{0\} \cup \left\{ \frac{1}{n} : n \in \mathbb{N} \right\}$$

with the induced metric d from the usual metric on $\mathbb{R}$. Then $\{0\}$ is a component of (X,d) but is not open in (X,d). In fact, if $C(0)$, the component of 0 in X, contains another point of X, say, $1/m$, then there exists an irrational number α such that $0 < \alpha < 1/m$. The sets $(-\infty,\alpha) \cap C(0)$, $(\alpha,\infty) \cap C(0)$ are nonempty, disjoint and open in $C(0)$, whose union is $C(0)$, a contradiction.

Theorem 6.15. *A connected subset A of a metric space X which is both open and closed is a component of X.*

Proof. Let C be a component of X. Then every connected subset of X is contained in C, that is, $A \subseteq C$. We claim that $A = C$. Suppose that $A \neq C$, then $C \cap A$ is a nonempty open set in C because A is open in X. Moreover, since A is a closed set in X, $X \setminus A$ is open and so $C \cap (X \setminus A)$ is a nonempty open set in C. Also, note that

$$[(C \cap A)] \cap [C \cap (X \setminus A)] = C \cap [A \cap (X \setminus A)] = C \cap \emptyset = \emptyset$$

and

$$[C \cap A] \cup [C \cap (X \setminus A)] = C \cap [A \cup (X \setminus A)] = C \cap X = C.$$

This shows that C is the union of two disjoint open sets $C \cap A$ and $C \cap (X \setminus A)$. Therefore, C is not connected which is a contradiction. Hence $A = C$.

Theorem 6.16. *Each point of a metric space X is contained in exactly one component of X.*

Proof. Let $x \in X$ be arbitrary. Consider the family $\{C_\alpha\}_{\alpha \in \Lambda}$ of all connected subsets of X such that $x \in C_\alpha$ for all $\alpha \in \Lambda$. Since the singleton set $\{x\}$ is connected, this family $\{C_\alpha\}_{\alpha \in \Lambda}$ is nonempty and $\bigcap_{\alpha \in \Lambda} C_\alpha \neq \emptyset$. By Theorem 6.8, $C = \bigcup_{\alpha \in \Lambda} C_\alpha$ is a connected subset of X which contains x. Clearly C is maximal, and therefore, a component of X. Finally, we show that C is the only component of X which contains x.

Suppose that C^* is another component of X containing x. Since C is the union of all connected subsets of X, $C^* \subseteq C$. Since C^* is maximal connected subset of X, we must have $C^* = C$.

Corollary 6.4. *Every connected subset of a metric space X is contained in a component of X.*

Exercise 6.13. Determine the components of a discrete metric space.

Exercise 6.14. If the number of components in a metric space is finite, then prove that each component is both open and closed.

Exercise 6.15. Let X and Y be metric spaces, C a component of Y and $f : X \to Y$ a continuous function. Prove that $f^{-1}(C)$ is the union of components of X.

Exercise 6.16. Let X be a compact metric space. If the components of X are open, then prove that there are only a finite number of components.

Exercise 6.17. Show that any pair of components are separated.

Chapter 7
Fixed Point Theorems

The Fixed Point Theory is divided into three major areas:

- Topological Fixed Point Theory
- Metric Fixed Point Theory
- Discrete Fixed Point Theory

The Banach contraction theorem is one of the most important and useful results in the metric fixed point theory. It is perhaps one of the most widely used fixed point theorems in all analysis. This is because the contraction condition on the mapping is simple and easy to verify, because it requires only completeness assumption on the underlying metric space, and because it finds almost canonical applications in the theory of differential and integral equations. Although the basic idea was known to others earlier, the theorem first appeared in explicit form in the setting of $C[0,1]$ in Banach's 1922 Ph.D. thesis where it was used to establish the existence of a solution of an integral equation. Caristi's fixed point theorem is also one of the most important results in metric fixed point theory because of its applications in different areas of science, engineering and social sciences.

This chapter deals with some important and useful results in metric fixed point theory. We present Banach contraction theorem and some of its applications. An important generalization of Banach contraction theorem, obtained by Boyd and Wong [18] in 1969, is given. We also present Caristi's fixed point theorem and some of its consequences.

7.1 Banach Contraction Theorem and its Applications

Definition 7.1. Let X be a nonempty set and $T : X \to X$ be a mapping. A point $\bar{x} \in X$ is said be a *fixed point* of T if $T(\bar{x}) = \bar{x}$.

Example 7.1. (a) Let $X = \mathbb{R}$ be a nonempty set and $T : X \to X$ be a mapping defined by $T(x) = x + a$ for some fixed number $a \neq 0$. Then T has no fixed point.

(b) Let $X = \mathbb{R}$ and $T : X \to X$ be a mapping defined by $T(x) = \frac{1}{2}x$. Then $x = 0$ is the only fixed point of T.

(c) Let $X = \mathbb{R}$ and $T : X \to X$ be a mapping defined by $T(x) = x^2$. Then $x = 0$ and $x = 1$ are two fixed points of T.

(d) Let $X = \mathbb{R}$ and $T : X \to X$ be a mapping defined by $T(x) = x$. Then T has infinitely many fixed points. In fact, every point of X is a fixed of T.

Remark 7.1. (i) Example 7.1 shows that a mapping may not have any fixed point, it may have a unique fixed point, it may have more than one fixed points or even infinitely many fixed points.

(ii) Let $X = \mathbb{R}$ be a metric space with the usual metric and $T : X \to X$ be a mapping defined by

$$T(x) = (1+x)^{1/3} \quad \text{for all } x \in X.$$

Then finding a fixed point of T is equivalent to solve the equation $x^3 - x - 1 = 0$.

Definition 7.2. Let (X,d) be a metric space. A mapping $T : X \to X$ is said to be *Lipschtiz continuous* if there exists a constant $\alpha > 0$ such that

$$d(T(x),T(y)) \leq \alpha d(x,y) \quad \text{for all } x,y \in X. \tag{7.1}$$

- If $\alpha = 1$, then T is said to be *nonexpansive* .
- If $\alpha \in (0,1)$, then T is said to be *contraction* .
- If $d(T(x),T(y)) < d(x,y)$ for all $x \neq y$, then T is said to be *contractive* .

The number α is called *Lipschitz constant* of T.

Remark 7.2. It is clear from the definition that every Lipschtiz continuous function is continuous and every contraction mapping is contractive. But converse assertions may not be true. For example, consider the metric space $X = [0,\infty)$ with the usual metric and the mapping $T : X \to X$ given by $T(x) = \frac{1}{1+x^2}$. Then T is contractive but not contraction.

Theorem 7.1 (Banach Contraction Theorem). *Let (X,d) be a complete metric space and $T : X \to X$ be a contraction mapping. Then T has a unique fixed point.*

Proof. We construct a sequence $\{x_n\}$ by the following iterative method.

Choose any arbitrary point $x_0 \in X$. Then $x_0 \neq T(x_0)$, otherwise x_0 is a fixed point of T and there is nothing to prove. So, we define

$$x_1 = T(x_0), x_2 = T(x_1), x_3 = T(x_2), \ldots, x_n = T(x_{n-1}) \quad \text{for all } n \in \mathbb{N}.$$

We claim that this sequence $\{x_n\}$ of points of X is a Cauchy sequence.

Since T is a contraction mapping with Lipschitz constant $0 < \alpha < 1$, for all $p = 1,2,\ldots$, we have

$$\begin{aligned} d(x_{p+1},x_p) &= d(T(x_p),T(x_{p-1})) \leq \alpha d(x_p,x_{p-1}) \\ &= \alpha d(T(x_{p-1}),T(x_{p-2})) \leq \alpha^2 d(x_{p-1},x_{p-2}) \\ &\quad \cdots\cdots\cdots\cdots \\ &\quad \cdots\cdots\cdots\cdots \\ &= \alpha^{p-1} d(T(x_1),T(x_0)) \leq \alpha^p d(x_1,x_0). \end{aligned}$$

Let m and n be any positive integers with $n < m$. Then by the triangle inequality, we have

$$\begin{aligned} d(x_m,x_n) &\leq d(x_m,x_{m-1}) + d(x_{m-1},x_{m-2}) + \cdots + d(x_{n+1},x_n) \\ &\leq \left(\alpha^{m-1} + \alpha^{m-2} + \cdots + \alpha^n\right) d(x_1,x_0) \\ &\leq \alpha^n \left(\alpha^{m-n-1} + \alpha^{m-n-2} + \cdots + 1\right) d(x_1,x_0) \\ &\leq \frac{\alpha^n}{1-\alpha} d(x_1,x_0). \end{aligned}$$

Since $\lim_{n\to\infty} \alpha^n = 0$ and $d(x_1,x_0)$ is fixed, the right hand side of the above inequality approaches to 0 as n tends to ∞. It follows that $\{x_n\}$ is a Cauchy sequence in X. Since X is complete, there exists $\bar{x} \in X$ such that $x_n \to \bar{x}$. We show that this limit point $\bar{x}$ is a fixed point of T.

Since T is a contraction mapping, from the triangle inequality, we have

$$\begin{aligned} d(\bar{x},T(\bar{x})) &\leq d(\bar{x},x_n) + d(x_n,T(\bar{x})) \\ &= d(\bar{x},x_n) + d(T(x_{n-1}),T(\bar{x})) \\ &\leq d(\bar{x},x_n) + \alpha d(x_{n-1},\bar{x}) \to 0 \text{ as } n \to \infty. \end{aligned}$$

Hence $d(\bar{x},T(\bar{x})) = 0$ and so $T(\bar{x}) = \bar{x}$.

Now we show that the fixed point of T is unique. Suppose to the contrary that x and y are two distinct fixed points of T. Then $T(x) = x$ and $T(y) = y$. Since T is a contraction mapping, we have

$$d(x,y) = d(T(x),T(y)) \leq \alpha d(x,y) < d(x,y)$$

a contradiction. Hence $x = y$.

Remark 7.3. (i) If X is not complete in Theorem 7.1, then T may not have a fixed point. For example, consider $X = (0,1)$ and the mapping $T : X \to X$ defined by $T(x) = \frac{x}{2}$. Then X is not a complete metric space with the usual metric and T does not have any fixed. In fact, $T(0) = 0 \notin X$.

(ii) If T is not contraction in Theorem 7.1, then it may not have a fixed point. For example, consider the metric space $X = [1,\infty)$ with the usual metric and the mapping $T : X \to X$ given by $T(x) = x + \frac{1}{x}$. Then X is a complete metric space but T is not a contraction mapping. In fact,

$$|T(x) - T(y)| = |x-y|\left(1 - \frac{1}{xy}\right) < |x-y| \quad \text{for all } x,y \in X$$

and so T is contractive. Of course, T does not have any fixed point.

The following example shows that if X is a complete metric space and $T : X \to X$ is not a contraction mapping but $T^2 = T \circ T$ is contraction, even then T has a fixed point.

Example 7.2. Let $X = \mathbb{R}$ be a metric space with the usual metric and $T : X \to X$ be a mapping defined as

$$T(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

Then f is not continuous and hence not a contraction mapping. But

$$T^2(x) = T(T(x)) = \begin{cases} T(1) = 1 & \text{if } x \text{ is rational} \\ T(0) = 1 & \text{if } x \text{ is irrational.} \end{cases}$$

Then T^2 is a contraction mapping but both T^2 and T have the same fixed point that is 1.

The above example motivates us to present the following result.

Theorem 7.2. *Let (X,d) be a complete metric space and $T : X \to X$ be a mapping such that for some integer m, $T^m = \underbrace{T \circ T \circ \cdots \circ T}_{m \text{ times}}$ is a contraction mapping. Then T has a unique fixed point.*

Proof. By Theorem 7.1, T^m has a unique fixed point $x \in X$, that is, $T^m(x) = x$. Then

$$T(x) = T(T^m(x)) = T^m(T(x))$$

and so $T(x)$ is a fixed point of T^m. Since the fixed point of T^m is unique, $T(x) = x$. To prove the uniqueness, we assume that y is another fixed point of T. Then $T(y) = y$ and so $T^m(y) = y$. Again by the uniqueness of the fixed point of T^m, we have $x = y$. Hence $x \in X$ is a unique fixed point of T.

The following theorem guarantees the existence of a fixed point for a contractive mapping.

Theorem 7.3. *Let (X,d) be a compact metric space and $T : X \to X$ be a contractive mapping. Then T has a unique fixed point $\bar{x}$. Moreover, for any $x \in X$, $\lim_{n\to\infty} T^n(x) = \bar{x}$, that is, the successive iterates $T(x), T^2(x), \ldots, T^n(x), \ldots$ converge to the unique fixed point $\bar{x}$ of T.*

Proof. Define a mapping $\psi : X \to \mathbb{R}_+$ by

$$\psi(x) = d(x, T(x)) \quad \text{for all } x \in X.$$

Then ψ is continuous. Indeed by contractiveness of T, we have

$$\begin{aligned} |\psi(x) - \psi(y)| &= |d(x,T(x)) - d(y,T(y))| \\ &\le |d(x,T(x)) - d(T(x),y)| + |d(T(x),y) - d(y,T(y))| \\ &\le d(x,y) + d(T(x),T(y)) \\ &< 2d(x,y). \end{aligned}$$

Let $\varepsilon > 0$ be given. Then

$$|\psi(x) - \psi(y)| < 2d(x,y) < \varepsilon \quad \text{whenever} \quad d(x,y) < \delta = \frac{\varepsilon}{2}.$$

Therefore, T is continuous. Clearly, ψ is bounded below. Since X is compact and ψ is continuous, there exists a minimizer $\bar{x} \in X$ of ψ, that is, there exists $\bar{x} \in X$ such that $\psi(\bar{x}) \le \psi(y)$ for all $y \in X$. We show that $\bar{x}$ is a fixed point of T.

Suppose contrary that $\bar{x}$ is not a fixed point of T. Then $T(\bar{x}) \neq \bar{x}$. By contractiveness of T, we have

$$\psi(T(\bar{x})) = d\left(T(\bar{x}), T^2(\bar{x})\right) < d(\bar{x}, T(\bar{x})) = \psi(\bar{x})$$

which contracts that $\bar{x}$ is a minimizer of ψ. Hence $\bar{x}$ is a fixed point of T. The uniqueness follows on the lines of the proof of Theorem 7.1.

Let $x \in X$. If $T^n(x) \neq \bar{x}$, then

$$d\left(T^{n+1}(x), \bar{x}\right) = d\left(T^{n+1}(x), T(\bar{x})\right) < d\left(T^n(x), \bar{x}\right) \quad \text{for all } n \in \mathbb{N}.$$

Therefore, $\{d(T^{n+1}(x),\bar{x})\}$ is a strictly decreasing sequence of nonnegative real numbers and so converges to its infimum. Suppose that $\alpha = \lim_{n\to\infty} d(T^{n+1}(x),\bar{x})$. Since $\{T^n(x)\}$ is a sequence of points of a compact metric space, there exists a subsequence $\{T^{n_k}(x)\}$ which converges to some point, say, $y \in X$. Since $\{T^n(x)\}$ is decreasing,

$$d(y,\bar{x}) = \lim_{k\to\infty} d(T^{n_k}(x),\bar{x}) = \lim_{k\to\infty} d(T^{n_k+1}(x),\bar{x}) = \alpha.$$

If $\alpha \neq 0$, then $y \neq \bar{x}$ and so we have

$$\begin{aligned}\alpha &= d(y,\bar{x}) > d(T(y),T(\bar{x})) = d(T(y),\bar{x})\\ &= \lim_{k\to\infty} d(T(T^{n_k}(x)),\bar{x}) = d(T^{n_k+1}(x),\bar{x})\\ &= \alpha\end{aligned}$$

a contradiction. Thus $\alpha = 0$, and therefore, $\lim_{n\to\infty} T^n(x) = \bar{x}$.

The following result is useful to test whether a differentiable function is contraction or not.

Theorem 7.4. *Let $X = [a,b]$ be a metric space with the usual metric and $T : [a,b] \to [a,b]$ be a differentiable function. Then T is a contraction mapping on X if and only if there exists a real number $\alpha < 1$ such that $|T'(x)| \leq \alpha$ for all $x \in (a,b)$.*

Proof. Let T be a contraction mapping. Then there exists $\alpha \in (0,1)$ such that

$$|T(x) - T(y)| \leq \alpha|x-y| \quad \text{for all } x,y \in [a,b].$$

Then for any x and $x+\Delta x$ in $[a,b]$, we have

$$|T(x+\Delta x) - T(x)| \leq \alpha\,|x+\Delta x - x| = \alpha|\Delta x|.$$

Hence, for $\Delta x \neq 0$,

$$\left|\frac{T(x+\Delta x)-T(x)}{\Delta x}\right| \leq \alpha.$$

Therefore,

$$|T'(x)| = \lim_{\Delta x\to 0}\left|\frac{T(x+\Delta x)-T(x)}{\Delta x}\right| \leq \alpha.$$

Conversely, assume that there exists a real number $\alpha < 1$ such that $|T'(x)| \leq \alpha$ for all $x \in (a,b)$. By Mean Value Theorem, for any $x \neq y$ in $[a,b]$, there exists $c \in (a,b)$ such that

$$T'(c) = \frac{T(x)-T(y)}{x-y}.$$

Since $|T'(c)| \leq \alpha$, we have

$$\left|\frac{T(x)-T(y)}{x-y}\right| = |T'(c)| \leq \alpha.$$

Hence

$$|T(x) - T(y)| \leq \alpha|x-y|$$

and therefore, T is a contraction mapping on $[a,b]$.

Exercise 7.1. Let $X=[1,\infty)$ be a metric space with the usual metric and $T:X\to X$ be a mapping defined by $T(x)=\frac{10}{11}\left(x+\frac{1}{x}\right)$ for all $x\in X$. Prove that T is a contraction mapping with Lipschtiz constant $\alpha=\frac{10}{11}$.

Exercise 7.2. Let $X=[0,1]$ be a metric space with the usual metric and $T:X\to X$ be a mapping defined by $T(x)=\frac{1}{7}\left(x^3+x^2+1\right)$ for all $x\in X$. Prove that T is a contraction mapping with Lipschtiz constant $\alpha=\frac{5}{7}$.
Hint: Use $\left|x^3-y^3\right|=|x-y|\ \left|x^2+xy+y^2\right|$ or use Theorem 7.4.

Exercise 7.3. Let $X=\{x\in\mathbb{Q}:x\geq 1\}$ be a metric space with the usual metric and $T:X\to X$ be a mapping defined by $T(x)=\frac{x}{2}+\frac{1}{x}$ for all $x\in X$. Prove that T is a contraction mapping with Lipschtiz constant $\alpha=\frac{1}{2}$.

Exercise 7.4. Let $X=\mathbb{R}$ be a metric space with the usual metric and $T:X\to X$ be a mapping defined by $T(x)=\cos x$ for all $x\in X$. Show that the mapping T is not contraction but contractive. Also show that the mapping $T(x)=\frac{99}{100}\cos x$ for all $x\in X$ is a contraction mapping on X.
Hint:

$$\begin{aligned}
|\cos x-\cos y| &= \left|2\sin\left(\frac{x-y}{2}\right)\sin\left(\frac{x+y}{2}\right)\right| \\
&= \left|2\sin\left(\frac{x-y}{2}\right)\right|\ \left|\sin\left(\frac{x+y}{2}\right)\right| \\
&\leq 2\left|\sin\left(\frac{x-y}{2}\right)\right| \\
&< 2\left|\frac{x-y}{2}\right| \quad (\text{since } |\sin w|<|w| \text{ for all } w\neq 0) \\
&= |x-y|.
\end{aligned}$$

Exercise 7.5. Let $X=\mathbb{R}$ be a metric space with the usual metric and $T:X\to X$ be a mapping defined by $T(x)=\sin x$ for all $x\in X$. Show that the mapping T is contractive.

Exercise 7.6. Let $X=\mathbb{R}^3$ be a metric space with the usual metric and $T:X\to X$ be a mapping defined by $T(x)=T(x_1,x_2,x_3)=\left(\frac{1}{2}\cos x_2+1,\frac{2}{3}\sin x_3,\frac{3}{4}x_1\right)$ for all $x=(x_1,x_2,x_3)\in X$. Show that the mapping T is a contraction on X and deduce that the simultaneous equations

$$\begin{aligned}
x_1 &= \frac{1}{2}\cos x_2+1, \\
x_2 &= \frac{2}{3}\sin x_3, \\
x_3 &= \frac{3}{4}x_1
\end{aligned}$$

have a unique solution.

Exercise 7.7. Let (X,d) be a complete metric space and $T:X\to X$ be a mapping such that

$$d(T(x),T(y)) \geq \alpha d(x,y) \quad \text{for some } \alpha > 1 \text{ and for all } x,y \in X.$$

Assume that $T(X) = X$. Prove that T is one-to-one. Moreover, prove that T has a unique fixed point $\bar{x} \in X$ with $\lim_{n\to\infty} T^{-n}(x) = \bar{x}$ for some $x \in X$.

7.1.1 Application to System of Linear Equations

In this subsection, we present an application of Banach contraction theorem to find the solution of the following system of n linear equations with n unknowns:

$$\left.\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n. \end{array}\right\} \tag{7.2}$$

This system can be written as

$$\left.\begin{array}{l} x_1 = (1-a_{11})x_1 - a_{12}x_2 - \cdots - a_{1n}x_n + b_1 \\ x_2 = -a_{21}x_1 + (1-a_{22})x_2 - a_{23}x_3 - \cdots - a_{2n}x_n + b_2 \\ x_3 = -a_{31}x_1 - a_{32}x_2 + (1-a_{33})x_3 - \cdots - a_{3n}x_n + b_3 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n = -a_{n1}x_1 - a_{n2}x_2 - a_{n3}x_3 - \cdots + (1-a_{nn})x_n + b_n. \end{array}\right\} \tag{7.3}$$

By letting $\alpha_{ij} = -a_{ij} + \delta_{ij}$, where

$$\delta_{ij} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}$$

system (7.3) is equivalent to the following system:

$$x_i = \sum_{j=1}^{n} \alpha_{ij}x_j + b_i \quad i = 1,2,3,\ldots,n. \tag{7.4}$$

If $x = (x_1,x_2,\ldots,x_n) \in \mathbb{R}^n$ and

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

and $b = (b_1,b_2,\ldots,b_n) \in \mathbb{R}^n$, then the system (7.4) is equivalent to

$$x = Ax + b. \tag{7.5}$$

In other words, the problem is to find the fixed point of the transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$T(x) = Ax + b. \tag{7.6}$$

If T is a contraction mapping, then we can use Banach Contraction Theorem 7.1 and obtain the unique solution of $T(x) = x$ by the method of successive approximation.

The conditions under which T is a contraction mapping depend on the choice of the metric on $X = \mathbb{R}^n$. Here we discuss one case and left two others for exercise.

Theorem 7.5. *Let $X = \mathbb{R}^n$ be a metric space with the metric $d_\infty(x,y) = \max_{1\le i\le n} |x_i - y_i|$. If*

$$\sum_{j=1}^{n} |\alpha_{ij}| \le \alpha < 1 \quad \textit{for all } i = 1,2,\ldots,n, \tag{7.7}$$

then the linear system (7.2) *of n linear equations in n unknowns has a unique solution.*

Proof. Since $X = \mathbb{R}^n$ with respect to the metric d_∞ is complete, it is sufficient to prove that the mapping T defined by (7.6) is a contraction.

$$\begin{aligned} d_\infty(T(x),T(y)) &= \max_{1\le i\le n} \left| \sum_{j=1}^{n} \alpha_{ij}(x_j - y_j) \right| \\ &\le \max_{1\le i\le n} \sum_{j=1}^{n} |\alpha_{ij}|\ |x_j - y_j| \\ &\le \max_{1\le i\le n} \left(\max_{1\le j\le n} |x_j - y_j| \right) \sum_{j=1}^{n} |\alpha_{ij}| \\ &= \max_{1\le i\le n} \sum_{j=1}^{n} |\alpha_{ij}|\, d_\infty(x,y) \\ &\le \alpha d_\infty(x,y). \end{aligned}$$

Thus T is a contraction mapping. By Banach Contraction Theorem 7.1, the linear systems (7.2) has a unique solution.

Exercise 7.8. Let $X = \mathbb{R}^n$ be a metric space with the metric $d_1(x,y) = \sum_{i=1}^{n} |x_i - y_i|$. If

$$\sum_{i=1}^{n} |\alpha_{ij}| \le \alpha < 1 \quad \text{for all } j = 1,2,\ldots,n, \tag{7.8}$$

then prove that the linear system (7.2) of n linear equations in n unknowns has a unique solution.

Hint:

$$d_1(T(x),T(y)) = \sum_{i=1}^{n} \left| \sum_{j=1}^{n} \alpha_{ij}(x_j - y_j) \right|$$

$$\begin{aligned}
&\leq \sum_{i=1}^{n}\sum_{j=1}^{n} |\alpha_{ij}| \ |x_j - y_j| \\
&= \sum_{j=1}^{n}\sum_{i=1}^{n} |\alpha_{ij}| \ |x_j - y_j| \\
&\leq \max_{1\leq j\leq n} \sum_{i=1}^{n} |\alpha_{ij}| \, d_1(x,y) \\
&\leq \alpha d_1(x,y).
\end{aligned}$$

Exercise 7.9. Let $X = \mathbb{R}^n$ be a metric space with the metric $d_2(x,y) = \left(\sum_{i=1}^{n} |x_i - y_i|^2\right)^{1/2}$. If

$$\sum_{i=1}^{n}\sum_{j=1}^{n} |\alpha_{ij}|^2 \leq \alpha^2 < 1, \tag{7.9}$$

then prove that the linear system (7.2) of n linear equations in n unknowns has a unique solution.
Hint:

$$\begin{aligned}
[d_2(T(x),T(y))]^2 &= \sum_{i=1}^{n} \left| \sum_{j=1}^{n} \alpha_{ij} (x_j - y_j) \right|^2 \\
&\leq \sum_{i=1}^{n} \left(\sum_{j=1}^{n} |\alpha_{ij}| \ |x_j - y_j| \right)^2 \\
&\leq \sum_{i=1}^{n} \left(\sum_{j=1}^{n} |\alpha_{ij}|^2 \sum_{j=1}^{n} |x_j - y_j|^2 \right).
\end{aligned}$$

So,

$$[d_2(T(x),T(y))]^2 \leq \sum_{i=1}^{n}\sum_{j=1}^{n} |\alpha_{ij}|^2 d_2(x,y) \leq \alpha^2 d_2(x,y).$$

7.1.2 Application to Differential Equations

We give an application of Banach contraction theorem to prove the existence and uniqueness of the following ordinary differential equation with an initial condition:

$$\frac{dy}{dx} = f(x,y), \quad y(x_0) = y_0. \tag{7.10}$$

Theorem 7.6 (Picard Theorem). *Let $f(x,y)$ be a continuous function of two variables defined on a rectangle $A = \{(x,y) : a \leq x \leq b, c \leq y \leq d\}$ and satisfy the following Lipschitz condition in the second variable:*

$$|f(x,y) - f(x,\hat{y})| \leq \alpha |y - \hat{y}| \quad \textit{for all } y,\hat{y} \in [c,d]. \tag{7.11}$$

Further, let (x_0,y_0) *be an interior point of* A. *Then the differential equation* (7.10) *with the given initial condition has a unique solution.*

Proof. First of all, we show that the problem of determining the solution of differential equation (7.10) is equivalent to the problem of finding the solution of an integral equation. If $y = g(x)$ satisfies the differential equation (7.10) and has the property that $g(x_0) = y_0$, then integrating differential equation (7.10) from x_0 to x, we obtain

$$\left.\begin{aligned} g(x) - g(x_0) &= \int_{x_0}^{x} f(t,g(t))dt \\ g(x) &= y_0 + \int_{x_0}^{x} f(t,g(t))dt. \end{aligned}\right\} \tag{7.12}$$

Thus a unique solution of the differential equation (7.10) with the given initial condition is equivalent to a unique solution of (7.12). We apply Banach contraction theorem to determine the solution of (7.12).

By (7.11), there exists a constant $q > 0$ such that

$$|f(x,y_1) - f(x,y_2)| \leq q|y_1 - y_2|.$$

Since $f(x,y)$ is continuous on a compact subset A of $\mathbb{R}^2$, it is bounded and so there exists a positive constant m such that $|f(x,y)| \leq m$ for all $(x,y) \in A$.

Choose a positive constant p such that $pq < 1$ and the rectangle $B = \{(x,y) : -p + x_0 \leq x \leq p + x_0,\ -pm + y_0 \leq y \leq pm + y_0\}$ is contained in A.

Let X be the set of all real-valued continuous functions $y = g(x)$ defined on $[-p + x_0, p + x_0]$ such that $d(g(x),y_0) \leq mp$. The set X is a closed subset of the metric space $C[x_0 - p, x_0 + p]$ with sup metric. Since $C[x_0 - p, x_0 + p]$ is complete, X is complete.

Consider the transformation $T : X \to X$ defined by

$$T(g) = h \quad \text{where} \quad h(x) = y_0 + \int_{x_0}^{x} f(t,g(t))dt.$$

Since

$$d(h(x),y_0) = \sup \left| \int_{x_0}^{x} f(t,g(t))dt \right| \leq m(x - x_0) \leq mp,$$

$h(x) \in X$, and so, T is well defined. For all $g, g_1 \in X$, we have

$$\begin{aligned} d(T(g,T(g_1)) = d(h,h_1) &= \sup \left| \int_{x_0}^{x} [f(t,g(t)) - f(t,g_1(t))]\,dt \right| \\ &\leq \int_{x_0}^{x} \sup |f(t,g(t)) - f(t,g_1(t))|\,dt \\ &\leq q \int_{x_0}^{x} |g(t) - g_1(t)|\,dt \\ &\leq qpd(g,g_1), \end{aligned}$$

that is,

$$d(T(g),T(g_1)) \leq \alpha d(g,g_1)$$

where $0 \leq \alpha = qp < 1$.

Hence T is a contraction mapping on X into itself. By Banach Contraction Theorem 7.1, T has a unique fixed point $g^* \in X$. This unique fixed point g^* is the unique solution of the differential equation (7.10) and satisfies the given initial condition.

7.1.3 Application to Integral Equations

The most interesting applications of fixed point theorems arise when the underlying metric space is a function space. Here we discuss the existence and uniqueness of the Volterra integral equation by using Theorem 7.3.

Let K be a continuous function on $[a,b] \times [a,b]$ and let ϕ be a continuous function on $[a,b]$. Consider the equation

$$f(x) = \phi(x) + \lambda \int_a^x K(x,y) f(y) dy \quad \text{for all } x \in [a,b], \tag{7.13}$$

where λ is a parameter. It is called the *Volterra equation.*

Theorem 7.7. *For each $\lambda \in \mathbb{R}$, the Volterra equation* (7.13) *has a unique solution f that is continuous on $[a,b]$.*

Proof. Let $X = C[a,b]$ the set of all continuous real-valued functions defined on $[a,b]$ with the uniform metric. Since K is continuous, there exists a constant $k > 0$ such that $|K(x,y)| \leq k$ for all $x,y \in [a,b]$. Define the transformation $T : f \mapsto T(f)$ on X by

$$T(f(x)) = \phi(x) + \lambda \int_a^x K(x,y) f(y) dy.$$

For all $f,g \in X$, we have

$$\begin{aligned} |T(f(x)) - T(g(x))| &= \left| \lambda \int_a^x K(x,y) |f(y) - g(y)| dy \right| \\ &\leq |\lambda| k (x-a) d(f,g) \quad \text{for all } x \in [a,b]. \end{aligned}$$

Since $T^2(f) - T^2(g) = T(T(f) - T(g))$, we have

$$\begin{aligned} |T^2(f(x)) - T^2(g(x))| &= \left| \lambda \int_a^x K(x,y) |T(f(y)) - T(g(y))| dy \right| \\ &\leq |\lambda| \int_a^x |K(x,y)| \, |\lambda| \, k \, (y-a) \, d(f,g) dy \\ &\leq |\lambda|^2 k^2 \int_a^x (y-a) dy \, d(f,g) \\ &\leq \frac{|\lambda|^2 k^2 (x-a)^2}{2} d(f,g). \end{aligned}$$

Continuing this iterative process, we obtain

$$|T^n(f(x)) - T^n(g(x))| \leq \frac{|\lambda|^n k^n (x-a)^n}{n!} d(f,g) \quad \text{for all } x \in [a,b].$$

Hence,

$$|T^n(f) - T^n(g)| \leq \frac{[|\lambda|k(b-a)]^n}{n!} d(f,g).$$

Recalling that $\frac{r^n}{n!} \to 0$ as $n \to \infty$ for any $r \in \mathbb{R}$, we conclude that there exists n such that T^n is a contraction mapping. Taking n sufficiently large to have $\frac{[|\lambda|k(b-a)]^n}{n!} < 1$. By Theorem 7.3, there exists a unique solution $f \in X$ satisfying $T(f) = f$. Obviously, if $T(f) = f$, then f solves (7.13).

7.2 Further Extension of Banach Contraction Theorem

Definition 7.3. Let X be a nonempty set. A function $f : X \to \mathbb{R}$ is said to be

- *bounded above* if there exists a real number k such that $f(x) \leq k$ for all $x \in X$;
- *bounded below* if there exists a real number k such that $k \leq f(x)$ for all $x \in X$;
- *bounded* if it is both bounded above as well as bounded below.

Definition 7.4. A function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ is said to be *upper semicontinuous from the right* if $r_n \downarrow r \geq 0$ implies $\limsup_{n\to\infty} \psi(r_n) \leq \psi(r)$.

In 1969, Boyd and Wong [18] obtained the following generalization of Banach contraction theorem.

Theorem 7.8. *Let (X,d) be a complete metric space and $\psi : [0,\infty) \to [0,\infty)$ be upper semicontinuous from the right such that $0 \leq \psi(t) < t$ for all $t > 0$. If $T : X \to X$ satisfies*

$$d(T(x),T(y)) \leq \psi(d(x,y)) \quad \text{for all } x,y \in X, \tag{7.14}$$

then it has a unique fixed point $\bar{x} \in X$ and $\{T^n(x)\}$ converges to $\bar{x}$ for all $x \in X$.

Proof. For any fixed $x \in X$, let $x_n = T^n(x)$, $n = 1,2,\ldots$ and $a_n = d(x_n,x_{n+1}) = d\left(T^n(x),T^{n+1}(x)\right)$. We show that a_n is convergent. We may assume that $a_n > 0$ for all $n > 0$. Then for all $n > 1$,

$$\begin{aligned} a_n &= d\left(T^n(x),T^{n+1}(x)\right) = d(T(x_{n-1}),T(x_n)) \\ &\leq \psi(d(x_{n-1},x_n)) = \psi(a_{n-1}) \\ &< a_{n-1}. \end{aligned}$$

Thus the sequence $\{a_n\}$ is monotonically decreasing and bounded below so it is convergent. Let $\lim_{n\to\infty} a_n = a$. We show that $a = 0$. If $a > 0$, then $a_{n+1} \leq \psi(a_n)$. By the upper semicontinuity from the right of the function ψ, we obtain $a \leq \psi(a)$ which is a contradiction with the property of ψ. Thus $a = 0$ and $a_n \to 0$ as $n \to \infty$.

We claim that $\{x_n\}$ is a Cauchy sequence. Assume to the contrary that the sequence $\{x_n\}$ is not Cauchy. Then there exists $\varepsilon > 0$ such that for any $k \in \mathbb{N}$, there exist $m_k > n_k \geq k$ such that

$$d\left(x_{m_k},x_{n_k}\right) \geq \varepsilon. \tag{7.15}$$

Furthermore, assume that for each k, m_k is the smallest number greater than n_k for which (7.15) holds. Let $a_k = d\left(x_{m_k}, x_{n_k}\right)$. Since $\lim_{n\to\infty} d(x_n, x_{n+1}) = \lim_{n\to\infty} a_n = 0$, there exists k_0 such that $d\left(x_k, x_{k+1}\right) \leq \varepsilon$ for all $k \geq k_0$. For such k, we have

$$\begin{aligned}\varepsilon &\leq d\left(x_{m_k}, x_{n_k}\right) \leq d\left(x_{m_k}, x_{m_k-1}\right) + d\left(x_{m_k-1}, x_{n_k}\right)\\ &\leq d\left(x_{m_k}, x_{m_k-1}\right) + \varepsilon \leq d\left(x_k, x_{k-1}\right) + \varepsilon.\end{aligned}$$

This proves $\lim_{k\to\infty} d\left(x_{m_k}, x_{n_k}\right) = \lim_{k\to\infty} a_k = \varepsilon$. On the other hand, we have

$$\begin{aligned}d\left(x_{m_k}, x_{n_k}\right) &\leq d\left(x_{m_k}, x_{m_k+1}\right) + d\left(x_{m_k+1}, x_{n_k+1}\right) + d\left(x_{n_k+1}, x_{n_k}\right)\\ &\leq a_{m_k} + \psi\left(d(x_{m_k}, x_{n_k})\right) + a_{n_k}\\ &\leq 2a_k + \psi\left(d(x_{m_k}, x_{n_k})\right).\end{aligned}$$

As $k \to \infty$, we obtain

$$\varepsilon = \lim_{k\to\infty} d\left(x_{m_k}, x_{n_k}\right) \leq \lim_{k\to\infty} \left(2a_k + \psi\left(d(x_{m_k}, x_{n_k})\right)\right) = \psi(\varepsilon).$$

Thus $\varepsilon \leq \psi(\varepsilon)$ which is a contradiction. Hence $\{T^n(x)\} = \{x_n\}$ is a Cauchy sequence.

Since $\{T^n(x)\}$ is a Cauchy sequence and X is complete, $\lim_{n\to\infty} T^n(x) = \bar{x} \in X$. Since T is continuous, $T(\bar{x}) = \bar{x}$. Uniqueness of $\bar{x}$ follows from condition (7.14).

Remark 7.4. If we replace the condition $\psi(t) < t$ by the condition $\psi(t_0) < t_0$ for at least one value to t_0, then Theorem 7.8 may fail. In this case, T may have no fixed point or else more than one fixed point.

Example 7.3. Let $X = (-\infty, -1] \cup [1, \infty)$ be a metric space with the absolute value metric. Let

$$T_1(x) = \begin{cases} \frac{1}{2}(x+1) & \text{if } x \geq 1 \\ \frac{1}{2}(x-1) & \text{if } x \leq -1 \end{cases}$$

and

$$T_2(x) = -T_1(x) \quad \text{for all } x \in X.$$

Then T_1 and T_2 satisfy condition (7.14) with

$$\psi(t) = \begin{cases} \frac{1}{2}t & \text{if } t < 2 \\ \frac{1}{2}t + 1 & \text{if } t \geq 2. \end{cases}$$

The function ψ satisfies all the conditions of Theorem 7.8 except that $\psi(2) = 2$. We observe that T_1 has two fixed points -1 and 1, while T_2 has no fixed points.

The following variant of the Boyd-Wong theorem is due to Matkowski [62]. In this result the continuity condition on ψ is replaced with another condition.

Theorem 7.9. *Let X be a complete metric space and $\psi : (0,\infty) \to (0,\infty)$ be monotone nondecreasing such that $\lim_{n\to\infty} \psi^n(t) = 0$ for all $t > 0$. If $T : X \to X$ satisfies*

$$d\left(T(x), T(y)\right) \leq \psi\left(d(x,y)\right) \quad \text{for all } x, y \in X,$$

then it has a unique fixed point $\bar{x}$ and $\lim_{n\to\infty} d\left(T^n(x),\bar{x}\right)=0$ *for all* $x\in X$.

Proof. For any fixed $x\in X$, let $x_n=T^n(x)$, $n=1,2,\ldots$. Then $x_1=T(x)\neq x$, otherwise, x would be a fixed point of T. Now

$$\begin{aligned} d\left(T^n(x),T^{n+1}(x)\right) &\leq \psi\left(d\left(T^{n-1}(x),T^n(x)\right)\right)\\ &\leq \psi^2\left(d\left(T^{n-2}(x),T^{n-1}(x)\right)\right)\\ &\vdots\\ &\leq \psi^n\left(d\left(x,T(x)\right)\right)=\psi^n\left(d\left(x,x_1\right)\right). \end{aligned}$$

Therefore,

$$0\leq \lim_{n\to\infty} d\left(x_n,x_{n+1}\right)=\lim_{n\to\infty} d\left(T^n(x),T^{n+1}(x)\right)\leq \lim_{n\to\infty}\psi^n\left(d\left(x,x_1\right)\right)=0.$$

Thus $\lim_{n\to\infty} d(x_n,x_{n+1})=0$. We show that $\{x_n\}$ is a Cauchy sequence.

Since $\psi^n(t)\to 0$ for all $t>0$, $\psi(\varepsilon)<\varepsilon$ for any $\varepsilon>0$. Since $\lim_{n\to\infty} d(x_n,x_{n+1})=0$, for any given $\varepsilon>0$, we can choose n such that

$$d\left(x_{n+1},x_n\right)\leq \varepsilon-\psi(\varepsilon).$$

Let $S_\varepsilon[x_n]=\{x\in X: d(x,x_n)\leq\varepsilon\}$. If $z\in S_\varepsilon(x_n)$, then $d(z,x_n)\leq\varepsilon$ and

$$\begin{aligned} d\left(T(z),x_n\right) &\leq d\left(T(z),T(x_n)\right)+d\left(T(x_n),x_n\right)\\ &\leq \psi\left(d\left(z,x_n\right)\right)+d\left(x_{n+1},x_n\right) \quad \text{as } T(x_n)=x_{n+1}\\ &\leq \psi(\varepsilon)+(\varepsilon-\psi(\varepsilon))=\varepsilon. \end{aligned}$$

Therefore, $T(z)\in S_\varepsilon[x_n]$ and so $T: S_\varepsilon[x_n]\to S_\varepsilon[x_n]$. It follows that $d\left(x_m,x_n\right)\leq\varepsilon$ for all $m\geq n$, and hence, $\{x_n\}$ is a Cauchy sequence.

The conclusion of the proof follows as in Theorem 7.8.

We now present a different kind of principle in which the contractive condition is imposed only at the first step.

Theorem 7.10. *Let X be a complete metric space and $T: X\to X$ be a continuous mapping such that for some function $\varphi: X\to[0,\infty)$, the following condition hold:*

$$d(x,T(x))\leq \varphi(x)-\varphi(T(x)) \quad \textit{for all } x\in X. \tag{7.16}$$

Then $\{T^n(x)\}$ converges to a fixed point of T for all $x\in X$.

Proof. For any fixed $x\in X$, let $x_n=T^n(x)$, $n=1,2,\ldots$. By the inequality (7.16), we have

$$0\leq \varphi(x)-\varphi(T(x)) \text{ if and only if } \varphi(T(x))\leq\varphi(x) \quad \text{for all } x\in X$$

and therefore,

$$\varphi(x_{n+1})=\varphi\left(T^{n+1}(x)\right)=\varphi\left(T(T^n(x))\right)=\varphi(T(x_n))\leq\varphi(x_n).$$

Thus, $\{\varphi(T^n(x))\} = \{\varphi(x_n)\}$ is monotonically decreasing and bounded below. Hence $\lim_{n\to\infty} \varphi(T^n(x)) = r \geq 0$. By the triangle inequality, if $m, n \in \mathbb{N}$ and $m > n$ then

$$\begin{aligned} d(T^n(x), T^m(x)) &\leq d\left(T^n(x), T^{n+1}(x)\right) + d\left(T^{n+1}(x), T^{n+2}(x)\right) \\ &\quad + \cdots + d\left(T^{m-1}(x), T^m(x)\right) \\ &\leq \varphi(T^n(x)) - \varphi\left(T^{n+1}(x)\right) + \varphi\left(T^{n+1}(x)\right) - \varphi\left(T^{n+2}(x)\right) \\ &\quad + \cdots + \varphi\left(T^{m-1}(x)\right) - \varphi(T^m(x)) \\ &\leq \varphi(T^n(x)) - \varphi(T^m(x)) \end{aligned}$$

so $\lim_{m,n\to\infty} d(T^n(x), T^m(x)) = 0$. It follows that $\{T^n(x)\} = \{x_n\}$ is a Cauchy sequence in X. Since X is complete, there exists $\bar{x} \in X$ such that $\lim_{n\to\infty} T^n(x) = \bar{x}$ and by continuity of T, $\bar{x} = T(\bar{x})$.

Remark 7.5. In Theorem 7.10, we can obtain an estimate on the rate of convergence of $\{T^n(x)\}$ by referring back to the inequality

$$\sum_{i=n}^{m-1} d\left(T^i(x), d(T^{i+1}(x)\right) \leq \varphi(T^n(x)) - \varphi(T^m(x)).$$

This yields

$$d(T^n(x), d(T^m(x)) \leq \varphi(T^n(x)) - \varphi(T^m(x)) \leq \varphi(T^n(x)),$$

and if $T(\bar{x}) = \bar{x}$, upon letting $m \to \infty$, we get

$$d(T^n(x), \bar{x}) \leq \varphi(T^n(x)).$$

Remark 7.6. If $T : X \to X$ is a contraction mapping then it is continuous and satisfies inequality (7.16).

Indeed, since T is a contraction mapping, we have

$$d\left(T(x), T^2(x)\right) \leq \alpha d(x, T(x)) \quad \text{for all } x \in X.$$

Adding $d(x, T(x))$ to both the sides of the above inequality yields

$$d(x, T(x)) + d\left(T(x), T^2(x)\right) \leq d(x, T(x)) + \alpha d(x, T(x)),$$

equivalently,

$$d(x, T(x)) - \alpha d(x, T(x)) \leq d(x, T(x)) - d\left(T(x), T^2(x)\right).$$

This in turn is equivalent to

$$d(x, T(x)) \leq \frac{1}{1-\alpha}\left[d(x, T(x)) - d\left(T(x), T^2(x)\right)\right].$$

Define the function $\varphi : X \to [0, \infty)$ by

$$\varphi(x) = \frac{1}{1-\alpha} d(x, T(x)) \quad \text{for all } x \in X.$$

This gives us the basic inequality

$$d(x, T(x)) \leq \varphi(x) - \varphi(T(x)) \quad \text{for all } x \in X.$$

7.3 Caristi's Fixed Point Theorem

Definition 7.5. Let X be a metric space. A function $\varphi : X \to \mathbb{R}$ is called

- *lower semicontinuous* at $x_0 \in X$ if

$$\liminf_{x \to x_0} \varphi(x) \geq \varphi(x_0);$$

- *upper semicontinuous* at $x_0 \in X$ if

$$\limsup_{x \to x_0} \varphi(x) \leq \varphi(x_0).$$

In general, φ is said to be lower semicontinuous (respectively, upper semicontinuous) on X if for any sequence $\{x_n\}$ in X such that $x_n \to x$ implies $\liminf_{n \to \infty} \varphi(x_n) \geq \varphi(x)$ $\left(\text{respectively, } \limsup_{n \to \infty} \varphi(x_n) \leq \varphi(x)\right)$.

Example 7.4. (a) Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a function defined by

$$\varphi(x) = \begin{cases} -1 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0. \end{cases}$$

Then φ is upper semicontinuous at $x_0 = 0$ but not lower semicontinuous.

(b) Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a function defined by

$$\varphi(x) = \begin{cases} -1 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0. \end{cases}$$

Then φ is lower semicontinuous at $x_0 = 0$ but not upper semicontinuous.

Theorem 7.11 (Caristi's Theorem). *Let (X, d) be a complete metric space and $T : X \to X$ be a mapping (not necessarily continuous). Suppose there exists a lower semicontinuous function $\varphi : X \to [0, \infty)$ such that*

$$d(x, T(x)) \leq \varphi(x) - \varphi(T(x)) \quad \textit{for all } x \in X. \tag{7.17}$$

Then T has a fixed point.

Proof. For any $x, y \in X$, define the relation $\preccurlyeq$ on X by

$$x \preccurlyeq y \quad \text{if and only if} \quad d(x, y) \leq \varphi(x) - \varphi(y).$$

It is easy to see that this ordering is (i) reflexive, that is, for all $x \in X$, $x \preccurlyeq x$; (ii) antisymmetric, that is, for all $x, y \in X$, $x \preccurlyeq y$ and $y \preccurlyeq x$ imply that $x = y$; (iii) transitive, that is, for all $x, y, z \in X$, $x \preccurlyeq y$ and $y \preccurlyeq z$ imply that $x \preccurlyeq z$. Then $(X, \preccurlyeq)$ is a partial ordered set. Let $x_0 \in X$ be an arbitrary but fixed element of X. Then by Zorn's Lemma B.1, we obtain a maximal totally ordered subset M of X containing x_0. Let $M = \{x_\alpha\}_{\alpha \in \Lambda}$, where Λ is totally ordered and

$$x_\alpha \preccurlyeq x_\beta \quad \text{if and only if} \quad \alpha \preccurlyeq \beta \quad \text{for all } \alpha, \beta \in \Lambda.$$

Since $\{\varphi(x_\alpha)\}_{\alpha \in \Lambda}$ is a decreasing net in $\mathbb{R}_+$, there exists $r \geq 0$ such that $\varphi(x_\alpha) \to r$ as α increases. Let $\varepsilon > 0$ be given. Then there exists $\alpha_0 \in \Lambda$ such that

$$\alpha \succcurlyeq \alpha_0 \quad \text{implies} \quad r \leq \varphi(x_\alpha) \leq r + \varepsilon.$$

Let $\beta \succcurlyeq \alpha \succcurlyeq \alpha_0$. Then

$$d(x_\alpha, x_\beta) \leq \varphi(x_\alpha) - \varphi(x_\beta) \leq r + \varepsilon - r = \varepsilon$$

which implies that $\{x_\alpha\}$ is a Cauchy net in X. Since X is complete, there exists $x \in X$ such that $x_\alpha \to x$ as α increases. From the lower semicontinuity of φ, we deduce that $\varphi(x) \leq r$. If $\beta \succcurlyeq \alpha$, then $d(x_\alpha, x_\beta) \leq \varphi(x_\alpha) - \varphi(x_\beta)$. Letting β increases, we obtain

$$d(x_\alpha, x) \leq \varphi(x_\alpha) - r \leq \varphi(x_\alpha) - \varphi(x)$$

which gives us $x_\alpha \preccurlyeq x$ for $\alpha \in \Lambda$. In particular, $x_0 \preccurlyeq x$. Since M is maximal, $x \in M$. Moreover, the condition (7.17) implies that

$$x_\alpha \preccurlyeq x \preccurlyeq T(x) \quad \text{for all } \alpha.$$

Again by maximality, $T(x) \in M$. Since $x \in M$, $T(x) \preccurlyeq x$ and hence $T(x) = x$.

Remark 7.7. As we have seen in Remark 7.6 that every contraction mapping $T : X \to X$ satisfies the condition 7.17 of the above theorem by choosing $\varphi(x) = \dfrac{1}{1-k} d(x, T(x))$.

Definition 7.6. A function $\xi : \mathbb{R} \to \mathbb{R}$ is said to be *concave downward* if for all $x, y \in \mathbb{R}$ and all $\lambda \in [0,1]$,

$$\xi(\lambda x + (1-\lambda)y) \geq \lambda \xi(x) + (1-\lambda)\xi(y).$$

Theorem 7.12. *Let (X,d) be a complete metric space and $\varphi : X \to \mathbb{R}$ be lower semicontinuous and bounded below. Suppose $\xi : \mathbb{R} \to \mathbb{R}$ is continuous, strictly increasing, concave downward, and vanishes at 0, and suppose $T : X \to X$ satisfies the following condition:*

$$\xi(d(x, T(x))) \leq \varphi(x) - \varphi(T(x)) \quad \textit{for all } x \in X.$$

Then T has a fixed point.

Proof. It is easy to see that $\xi \circ d$ is a metric on X. Suppose that $\{x_n\}$ is a Cauchy sequence in $(X, \xi \circ d)$. Since ξ is strictly increasing and vanishes at 0, $\lim_{m,n \to \infty} \xi(d(x_n, x_m))$

$= 0$ implies $\lim_{m,n\to\infty} d(x_n,x_m) = 0$. Since X is complete, there exists $x \in X$ such that $\lim_{n\to\infty} d(x_n,x) = 0$. By continuity of ξ this in turn implies $\lim_{n\to\infty} \xi(d(x_n,x)) = 0$. Therefore, $\{x_n\}$ converges to x in $(X, \xi \circ d)$, proving that $(X, \xi \circ d)$ is itself complete. Now we can apply Caristi's theorem directly to the space $(X, \xi \circ d)$ to obtain a fixed point x for T, completing the proof.

Corollary 7.1. *Let (X,d) be a complete metric space and $\varphi : X \to \mathbb{R}$ be lower semicontinuous and bounded below. Suppose that for some $r \in (0,1]$, $T : X \to X$ satisfies the following condition:*

$$d(x,T(x))^r \leq \varphi(x) - \varphi(T(x)) \quad \textit{for all } x \in X.$$

Then T has a fixed point.

Proof. Take $\xi(t) = t^r$, $t \in \mathbb{R}$, and apply the previous theorem.

Remark 7.8. It is known [8] that Corollary 7.1 is false if $r > 1$.

Chapter 8
Set-Valued Maps

The first natural instance when set-valued maps occur is the inverse f^{-1} of a single-valued map f from a set X to another set Y. We can always define f^{-1} as a set-valued map which associates with any data y the (possibly empty) set of solutions

$$f^{-1}(y) = \{x \in X : f(x) = y\}$$

to the equation $f(x) = y$. Kuratowski realized the importance of set-valued maps, also called multivalued maps or point-to-set maps or multifunctions, and devoted considerable space in his famous book on Topology. Other eminent mathematicians, namely, Painlevé, Hausdorff and Bouligand have also visualized the vital role of set-valued maps as one often encounter with such objects in concrete and real life problems. However, the authors of Bourbaki's volume TOPOLOGIE GÉNERALE emphasized to study properties of set-valued maps as single-valued maps from a set to the power set of another set, or factorizing single-valued maps to make them bijective. This came as a stumbling block in the progress of this field. The set-valued maps started popular among mathematicians working in the areas of game theory and economics when Kakutani fixed point theorem [55], an extension of Brouwer fixed point theorem to set-valued maps, came in existence. It was motivated by Von Neumann. In 1969, Nadler [64] extended Banach fixed point theorem to set-valued maps which plays an important and vital role in the theory of variational inequalities, differential inclusions, control theory, fractal geometry, etc. Since than, the theory of set-valued maps became an important subject for mathematicians working in nonlinear analysis, optimization theory, game theory, control theory, economics, fractal geometry, etc.

In view of such a wide variety of applications, most of the basic notions and results of single-valued maps have been extended to set-valued maps. These includes:

- Limits and continuity
- Nonlinear Functional Analysis (existence and approximation of solutions to equations and inclusions; fixed point theory; etc)
- Tangents and Normals
- Differentiation of Set-valued Maps
- Convergence of Set-valued Maps
- Measures and Integration
- Differential Inclusions

In this chapter, we focus our attention on the basic properties, continuity and fixed point theory for set-valued maps. For applications of set-valued maps to fractal geometry, see [9]. For other topics mentioned above and the applications of set-valued maps, we refer to [5, 7, 9, 13, 21, 46, 51, 57, 61] and the references therein.

8.1 Basic Concepts and Definitions

Definition 8.1. Let X and Y be two nonempty sets. A *set-valued map* or *multivalued map* or *point-to-set map* or *multifunction* $F : X \rightrightarrows Y$ from X to Y is a map that associates with any $x \in X$ a subset $F(x)$ of Y; the set $F(x)$ is called the *image of* x under F. F is called *proper* if there exists at least an element $x \in X$ such that $F(x) \neq \emptyset$, that is, F is not the constant map $\emptyset$. In this case the set $Dom(F) = \{x \in X : F(x) \neq \emptyset\}$ is called the *domain* of F. Actually, a set-valued map F is characterized by its *graph*, the subset of $X \times Y$ defined by

$$Graph(F) = \{(x,y) : y \in F(x)\}.$$

Indeed, if A is a nonempty subset of the product space $X \times Y$, then the graph of a set-valued map F is defined by

$$y \in F(x) \quad \text{if and only if} \quad (x,y) \in A.$$

The domain of F is the projection of $Graph(F)$ on X and the *image* of F, the subset of Y defined by

$$Im(F) = \bigcup_{x \in X} F(x) = \bigcup_{x \in Dom(F)} F(x),$$

is projection of $Graph(F)$ on Y. A set-valued map F from X to Y is called *strict* if $Dom(F) = X$, that is, if the image $F(x)$ is nonempty for all $x \in X$. Let K be a nonempty subset of X and let F be a strict set-valued map from K to Y. It may be useful to extend it to the set-valued map F_K from X to Y defined by

$$F_K(x) = \begin{cases} F(x) & \text{when } x \in K \\ \emptyset & \text{when } x \notin K \end{cases}$$

whose domain $Dom(F_K)$ is K.

When F is a set-valued map from X to Y and $K \subseteq X$, we denote by $F_{|_K}$ its restriction to K.

Example 8.1. (a) Let $X = \mathbb{R}$ and $F : X \rightrightarrows X$ be defined as

$$F(x) = \begin{cases} \{1\} & \text{if } x < 0 \\ \{0,1\} & \text{if } x = 0 \\ \{0\} & \text{if } x > 0. \end{cases}$$

Then F is a set-valued map.

(b) Let $X = \mathbb{R}$ and let $F : X \rightrightarrows X$ be defined as

$$F(x) = \begin{cases} \{1\} & \text{if } x < 0 \\ [0,1] & \text{if } x = 0 \\ \{0\} & \text{if } x > 0. \end{cases}$$

Then F is a set-valued map.

(c) Let $X = [0,1]$ and let $F : X \rightrightarrows X$ be defined as

$$F(x) = \begin{cases} [0, \frac{1}{2}] & \text{if } x \neq \frac{1}{2} \\ [0,1] & \text{if } x = \frac{1}{2}. \end{cases}$$

Then F is a set-valued map.

(d) Let $X = \mathbb{R}_+$, $Y = \mathbb{R}$ and $F : X \rightrightarrows Y$ be defined by

$$F(x) = [e^{-x}, 1] \quad \text{for all } x \in X.$$

Then F is a set-valued map.

(e) Let $I^2 = \{(x,y) : 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1\}$ and let $F : I^2 \rightrightarrows I^2$ be defined as follows:

$F(x,y)$ is the line segment in I^2 from the point $(x,0)$ to the point $(0,y)$ for each $(x,y) \in I^2$.

Then F is a set-valued map.

Example 8.2 (Inverse Function). If $f : X \to Y$ is a single-valued map then it inverse f^{-1} can be considered as a set-valued map $F : Y \to X$ defined by

$$F(y) = f^{-1}(y) \quad \text{for all } y \in Im(f).$$

This set-valued map is strict when f is surjective, and single-valued when f is injective. This map plays an important role when we study the equation $f(x) = y$ and the behaviour of the set of solutions of $f^{-1}(y)$ as y ranges over Y.

Example 8.3 (Metric Projection). Let (X,d) be a metric space and K be a nonempty compact subset of X. Then by Exercise 5.9, for every $x \in X$, there exists a $y \in K$ such that $d(x,y) = \rho(x,K)$, where $\rho(x,K) = \inf_{z \in K} d(x,z)$ is the distance from a point x to the set K.

The *metric projection* is a set-valued map $P : X \rightrightarrows K$ defined by

$$P(x) = \{y \in K : d(x,y) = \rho(x,K)\} \quad \text{for all } x \in X.$$

Definition 8.2. Let $F : X \rightrightarrows Y$ be a set-valued map. For a nonempty subset A of X, we write

$$F(A) = \bigcup_{x \in A} F(x).$$

If $A = \emptyset$, we write $F(\emptyset) = \emptyset$. The set $F(A)$ is called the *image* of A under the set-valued map F.

Definition 8.3. Let $F : X \rightrightarrows Y$ be a set-valued map.

- If $F(X) = Y$, then F is called *quasi-surjective.*
- If $F(x_1) \cap F(x_2) \neq \emptyset$ implies $F(x_1) = F(x_2)$ for any $x_1, x_2 \in X$, then F called *hyperunivocal or semi-single-valued map.*
- If $F(x_1) \cap F(x_2) \neq \emptyset$ implies $x_1 = x_2$ for any $x_1, x_2 \in X$, then F called *hyperinjective.*

Example 8.4. (a) Let F be a set-valued map from $\mathbb{R}^3$ to itself and O be a fixed point in $\mathbb{R}^3$. Given $x \in \mathbb{R}^3$, let $F(x)$ be the set of points other then O on the straight line joining x to O. This determines a hyperunivocal set-valued map F whose domain is $\mathbb{R}^3 - \{O\}$.

(b) Let F be a set-valued map from $\mathbb{R}$ to itself defined as

$$F(x) = \begin{cases} \{\frac{1}{x}\} & \text{if } x \neq 0 \\ \{0\} & \text{if } x = 0. \end{cases}$$

Then F is hyperinjective.

Remark 8.1. (i) Every single-valued map is semi-single-valued.
(ii) If $x_1 \neq x_2$ implies $F(x_1) \cap F(x_2) = \emptyset$, then F is hyperinjective.

Theorem 8.1. *Let $\{A_\alpha\}_{\alpha \in \Lambda}$ be a family of nonempty subsets of X and $F : X \rightrightarrows Y$ be a set-valued map.*

(a) If $A_1 \subseteq A_2$ then $F(A_1) \subseteq F(A_2)$;

(b) $F\left(\bigcup_{\alpha \in \Lambda} A_\alpha\right) = \bigcup_{\alpha \in \Lambda} F(A_\alpha)$;

(c) $F\left(\bigcap_{\alpha \in \Lambda} A_\alpha\right) \subset \bigcap_{\alpha \in \Lambda} F(A_\alpha)$;

(d) $F(X \setminus A_1) \supseteq F(X) \setminus F(A_1)$. Further, if F is hyperinjective then $F(X \setminus A_1) = F(X) \setminus F(A_1)$.

Proof. (a) If $y \in F(A_1)$, then $y \in F(x)$ for some $x \in A_1$. Since $A_1 \subseteq A_2$, we have $y \in F(x)$ for some $x \in A_2$, whence $y \in F(A_2)$.

(b) If $y \in F(\bigcup_{\alpha \in \Lambda} A_\alpha)$, then $y \in F(x)$ for some $x \in \bigcup_{\alpha \in \Lambda} A_\alpha$, whence, for at least one index α_0, we have $y \in F(A_{\alpha_0})$. Therefore, $y \in \bigcup_{\alpha \in \Lambda} F(A_\alpha)$.

By reversing the argument, we can deduce the desired formula.

(c) If $y \in F(\bigcap_{\alpha \in \Lambda} A_\alpha)$, then $y \in F(x)$ for some x such that $x \in A_\alpha$ for all $\alpha \in \Lambda$, and so $y \in F(A_\alpha)$ for all $\alpha \in \Lambda$, whence $y \in \bigcap_{\alpha \in \Lambda} F(A_\alpha)$ and so the result is proved.

(d) Let $y \in F(X) \setminus F(A_1)$. Then $y \in \bigcup_{x \in X} F(x) \setminus \bigcup_{x \in A_1} F(x)$. Therefore, $y \in F(x)$ for some $x \in X$ but $x \notin A_1$. This implies that $y \in F(x)$ for some $x \in X \setminus A_1$ and hence $y \in \bigcup_{x \in X \setminus A_1} F(x)$. Therefore, $y \in F(X \setminus A_1)$, and so $F(X) \setminus F(A_1) \subseteq F(X \setminus A_1)$.

If F is hyperinjective, then $y \in F(X \setminus A_1)$ implies that there exists $x \in X \setminus A_1$ such that $y \in F(x)$. Hence $y \notin F(A_1)$ and so $y \in F(X) \setminus F(A_1)$. Thus $F(X \setminus A_1) = F(X) \setminus F(A_1)$.

Definition 8.4. If F_1 and F_2 are two set-valued maps from X to Y.

- The *union* of F_1 and F_2 is a set-valued map $(F_1 \cup F_2)$ from X to Y defined by

$$(F_1 \cup F_2)(x) = F_1(x) \cup F_2(x) \quad \text{for all } x \in X.$$

- The *intersection* of F_1 and F_2 is a set-valued map $(F_1 \cap F_2)$ from X to Y defined by

$$(F_1 \cap F_2)(x) = F_1(x) \cap F_2(x) \quad \text{for all } x \in X.$$

- The *Cartesian product* of F_1 and F_2 is a set-valued map $(F_1 \times F_2)$ from X to $Y \times Y$ defined by

$$(F_1 \times F_2)(x) = F_1(x) \times F_2(x) \quad \text{for all } x \in X.$$

- If F_1 is a set-valued map from X to Y and F_2 is another set-valued map from Y to Z, then the *composition product* of F_2 by F_1 is a set-valued map $(F_2 \circ F_1)$ from X to Z defined by

$$(F_2 \circ F_1)(x) = F_2(F_1(x)).$$

Theorem 8.2. *If F_1 and F_2 are two set-valued maps from X to Y and A is a nonempty subset of X. Then*

(a) $(F_1 \cup F_2)(A) = F_1(A) \cup F_2(A)$;
(b) $(F_1 \cap F_2)(A) \subseteq F_1(A) \cap F_2(A)$;
(c) $(F_1 \times F_2)(A) \subseteq F_1(A) \times F_2(A)$;
(d) $(F_2 \circ F_1)(A) = F_2(F_1(A))$.

Theorem 8.3. *If F_1 and F_2 are two hyperunivocal set-valued maps from X to Y, then $F_1 \cap F_2$ and $F_1 \times F_2$ are also hyperunivocal.*

Proof. We have

$$\Big((F_1 \cap F_2)(x_1)\Big) \cap \Big((F_1 \cap F_2)(x_2)\Big) \neq \emptyset \quad \text{for all } x_1, x_2 \in X.$$

This implies that

$$\begin{cases} F_1(x_1) \cap F_1(x_2) \neq \emptyset \\ F_2(x_1) \cap F_2(x_2) \neq \emptyset. \end{cases}$$

Since F_1 and F_2 are hyperunivocal, we have

$$F_1(x_1) = F_1(x_2) \quad \text{and} \quad F_2(x_1) = F_2(x_2).$$

Thus $F_1(x_1) \cap F_2(x_1) = F_1(x_2) \cap F_2(x_2)$ and hence $(F_1 \cap F_2)(x_1) = (F_1 \cap F_2)(x_2)$.

The proof for the Cartesian product $F_1 \times F_2$ is similar.

Theorem 8.4. *If one of the set-valued maps F_1 and F_2 from X to Y is hyperinjective then the map $F_1 \cap F_2$ and $F_1 \times F_2$ are hyperinjective.*

Proof. If x and y are two distinct elements of X, then

$$\Big(F_1(x) \cap F_2(x)\Big) \cap \Big(F_1(y) \cap F_2(y)\Big)$$

$$= \Big(F_1(x) \cap F_1(y)\Big) \cap \Big(F_2(x) \cap F_2(y)\Big) = \emptyset.$$

The proof for the Cartesian product $F_1 \times F_2$ is similar.

Definition 8.5. If F is a set-valued map from X to Y, then the *inverse* F^{-1} of F is defined by

$$F^{-1}(y) = \{x \in X : y \in F(x)\} \quad \text{for all } y \in Y$$

Further, let B be a subset of Y. The *upper inverse image* $F^{-1}(B)$ and *lower inverse image* $F_+^{-1}(B)$ of B under F are defined by

$$F^{-1}(B) = \{x \in X : F(x) \cap B \neq \emptyset\}$$

$$F_+^{-1}(B) = \{x \in X : F(x) \subseteq B\}.$$

We also write $F^{-1}(\emptyset) = \emptyset$ and $F_+^{-1}(\emptyset) = \emptyset$. It is clear that the definition of inverse of F^{-1} that $(F^{-1})^{-1} = F$ and that $y \in F(x)$ if and only if $x \in F^{-1}(y)$.

We have the following relations between domain, graphs and images of F and F^{-1}.

$$Dom(F^{-1}) = Im(F), \qquad Im(F^{-1}) = Dom(F) \text{ and}$$

$$Graph(F^{-1}) = \{(y,x) \in Y \times X : (x,y) \in Graph(F)\}.$$

Example 8.5. (a) Let $X = Y = \mathbb{R}_+$ and $F : X \rightrightarrows Y$ be a set-valued map defined by

$$F(x) = [0,x] \quad \text{for all } x \in X.$$

Since $F(\mathbb{R}_+) = \mathbb{R}_+$, the map F^{-1} is defined on all the nonnegative half-axis. For any $y \in \mathbb{R}_+$,

$$F^{-1}(y) = \{x \in X : y \in [0,x]\} = [y, +\infty).$$

(b) Let F be a set-valued map from $[0,1]$ to $[0,2]$ defined by

$$F(x) = [0,x+1] \quad \text{for all } x \in [0,1].$$

Then $F([0,1]) = \bigcup_{x \in [0,1]} F(x) = [0,2]$ and hence F is quasisurjective.

$$F^{-1}([0,3/2]) = \{x \in [0,1] : F(x) \cap [0,3/2] \neq \emptyset\} = [0,1].$$

$$F^{-1}\left(\frac{1}{2}\right) = \{x \in [0,1] : 1/2 \in F(x)\} = [0,1].$$

Theorem 8.5. *(a) If $F : X \to Y$ is single-valued, then $F^{-1} : Y \rightrightarrows X$ is hyperinjective.*
(b) If the set-valued map $F : X \rightrightarrows Y$ is hyperinjective, then F^{-1} is single-valued.
(c) If the set-valued map $F : X \rightrightarrows Y$ is hyperunivocal, then F^{-1} is also hyperunivocal.

Proof. (a) If F is single-valued, then

$$x \neq y \quad \text{implies} \quad F^{-1}(x) \cap F^{-1}(y) = \emptyset,$$

and whence F^{-1} is hyperinjective.

(b) If F is hyperinjective, then the set $F^{-1}(y) = \{x \in X : y \in F(x)\}$ has only one element, and therefore, F^{-1} is single-valued.

(c) If F is hyperunivocal, then we have $F^{-1}(x) \cap F^{-1}(y) \neq \emptyset$. Therefore, there exists $z \in X$ such that $x \in F(z)$ and $y \in F(z)$. Thus $F^{-1}(x) = F^{-1}(y)$ and hence F^{-1} is hyperunivocal.

Theorem 8.6. *Let $\{B_\alpha\}_{\alpha \in \Lambda}$ be a family of nonempty subsets of Y, $A \subseteq X$ and $B \subseteq Y$. Let $F : X \rightrightarrows Y$ be a set-valued map.*

(a) If $B_1 \subseteq B_2$ then $F^{-1}(B_1) \subseteq F^{-1}(B_2)$;
(b) $A \subset F_+^{-1}(F(A))$;

(c) $B \subset F(F_+^{-1}(B))$;
(d) $X \setminus F_+^{-1}(B) = F_+^{-1}(Y \setminus B)$;
(e) $F_+^{-1}\left(\bigcup_{\alpha\in\Lambda} B_\alpha\right) \subset \bigcup_{\alpha\in\Lambda} F_+^{-1}(B_\alpha)$;
(f) $F_+^{-1}\left(\bigcap_{\alpha\in\Lambda} B_\alpha\right) = \bigcap_{\alpha\in\Lambda} F_+^{-1}(B_\alpha)$;
(g) $F^{-1}(F(A)) \subset A$;
(h) $F(F^{-1}(B)) \subset B \cap F(X)$;
(i) $X \setminus F^{-1}(B) = F_+^{-1}(Y \setminus B)$;
(j) $F^{-1}\left(\bigcap_{\alpha\in\Lambda} B_\alpha\right) \subset \bigcap_{\alpha\in\Lambda} F^{-1}(B_\alpha)$;
(k) $F^{-1}\left(\bigcup_{\alpha\in\Lambda} B_\alpha\right) = \bigcup_{\alpha\in\Lambda} F^{-1}(B_\alpha)$.

Theorem 8.7. *Let $F_1, F_2 : X \to Y$ be set-valued maps such that $(F_1 \cap F_2)(x) \neq \emptyset$ for all $x \in X$ and let $B \subseteq Y$. Then*

(a) $(F_1 \cup F_2)^{-1}(B) = F_1^{-1}(B) \cup F_2^{-1}(B)$;
(b) $(F_1 \cap F_2)^{-1}(B) \subset F_1^{-1}(B) \cap F_2^{-1}(B)$;
(c) $(F_1 \cup F_2)_+^{-1}(B) = F_{1+}^{-1}(B) \cap F_{2+}^{-1}(B)$;
(d) $(F_1 \cap F_2)_+^{-1}(B) \subset F_{1+}^{-1}(B) \cup F_{2+}^{-1}(B)$;

Theorem 8.8. *Let $F_1 : X \to Y$ and $F_2 : Y \to Z$ be set-valued maps. Then for any $B \subseteq Z$ we have*

(a) $(F_2 \circ F_1)_+^{-1}(B) = F_{1+}^{-1}\left(F_{2+}^{-1}(B)\right)$;
(b) $(F_2 \circ F_1)^{-1}(B) = F_1^{-1}\left(F_2^{-1}(B)\right)$.

Theorem 8.9. *Let $F_1 : X \to Y$ and $F_2 : X \to Z$ be set-valued maps. Then for any $B \subseteq Y$ and $D \subseteq Z$ we have*

(a) $(F_1 \times F_2)_+^{-1}(B \times D) = F_{1+}^{-1}(B) \cap F_{2+}^{-1}(D)$;
(b) $(F_1 \times F_2)^{-1}(B \times D) = F_1^{-1}(B) \cap F_2^{-1}(D)$.

8.2 Continuity of Set-Valued Maps

Let (X,d) be a metric space and K be a subset of X. Then the distance from a point $x \in X$ to the set K is defined by

$$d(x,K) = \inf_{y\in K} d(x,y).$$

The *sphere of radius $r > 0$ around K* is denoted by

$$S_r[K] = \{x \in X : d(x,K) \leq r\}.$$

The sphere $S_r[K]$ is a neighborhood of K. If K is compact, then $S_r[K]$ is closed.

Indeed, if K is compact then it is totally bounded. Therefore, the set $S_r[K]$ can be expressed as the union of a finite number of closed sets. Thus, it is closed.

When K is compact, each neighborhood of K contains such a sphere around K.

If the images of a set-valued map F are closed, bounded, compact and so on, we say that F is *closed valued, bounded valued, compact valued*, and so on.

Recall that for single-valued maps, continuity amounts to mapping converging sequences to converging sequences or, equivalently, to the celebrated incantation:

$$\text{for any } \varepsilon > 0, \text{ there exists } \delta > 0, \ldots\ldots$$

Unfortunately, this characterization is no longer true for set-valued maps.

There are two distinct ways to extend the concept of continuity: The first one, encompassing the above "convergence" idea, is called *lower semicontinuous*. The second, extending the "(for any $\varepsilon > 0$, there exists $\delta > 0, \ldots\ldots$) definition", leads to the so-called *upper semicontinuous* of set-valued maps. The concept of two kinds of semicontinuity of a set-valued map was introduced by G. Bouligand [17] and K. Kuratowski [59].

Definition 8.6. Let X and Y be metric spaces. A set-valued map $F : X \rightrightarrows Y$ is called *upper semicontinuous* at $x_0 \in Dom(F)$ if for any neighborhood U of $F(x_0)$, there exists $\delta > 0$ such that $F(x) \subseteq U$ for all $x \in S_\delta(x_0)$.

It is said to be *upper semicontinuous* on X if it is upper semicontinuous at every point of $Dom(F)$.

When $F(x)$ is compact for all $x \in X$, F is upper semicontinuous at x_0 if and only if for any $\varepsilon > 0$, there exists $\delta > 0$ such that $F(x) \subseteq S_\varepsilon(F(x))$ for all $x \in S_\delta(x_0)$

We observe that this definition is a natural adaptation of the definition of a continuous single-valued map. Why then do we use the adjective upper semicontinuous instead of continuous? One of the reason is that the celebrated characterization of continuous maps – "a single-valued map f is continuous at x if and only if it maps sequences converging to x to sequences converging to $f(x)$" – does not hold true any longer in the set-valued case. Indeed, the set-valued version of this characterization leads to the following definition.

Definition 8.7. Let X and Y be metric spaces. A set-valued map $F : X \rightrightarrows Y$ is called *lower semicontinuous* at $x_0 \in Dom(F)$ if for any $y \in F(x_0)$ and for any sequence of elements $x_n \in Dom(F)$ converging to x_0, there exists a sequence of elements $y_n \in F(x_n)$ converging to y.

It is said to be *lower semicontinuous* on X if it is lower semicontinuous at every point $x \in Dom(F)$.

The above definition of lower semicontinuity could be interpreted as follows:

F is *lower semicontinuous* at x_0 if for any $y_0 \in F(x_0)$ and any neighborhood $V(y_0)$ of y_0, there exists $\delta > 0$ such that $F(x) \cap V(y_0) \neq \emptyset$ for all $x \in S_\delta(x_0)$.

Actually, as in the single-valued case, this definition is equivalent to the following definition.

For any open subset $U \subseteq Y$ such that $U \cap F(x_0) \neq \emptyset$, there exists $\delta > 0$ such that $F(x) \cap U \neq \emptyset$ for all $x \in S_\delta(x_0)$.

When $F(x)$ is compact, F is lower semicontinuous at x_0 if and only if for any $\varepsilon > 0$, there exists $\delta > 0$ such that $F(x) \cap S_\varepsilon(F(x)) \neq \emptyset$ for all $x \in S_\delta(x_0)$.

The following examples show that the notions of upper semicontinuity and lower semicontinuity are not equivalent.

Example 8.6. (a) Let $X = [0,1]$ be a metric space with the usual metric and $F : X \rightrightarrows X$ be a set-valued map defined by

$$F(x) = \begin{cases} \frac{1}{2}x & \text{if } 0 \leq x < \frac{1}{2} \\ [\frac{1}{4}, \frac{3}{4}] & \text{if } x = \frac{1}{2} \\ \frac{1}{2}(x+1) & \text{if } \frac{1}{2} < x < 1. \end{cases}$$

Then F is upper semicontinuous but not lower semicontinuous.

(b) Let $X = \mathbb{R}$ be a metric space with the usual metric and $F : X \rightrightarrows X$ be a set-valued map defined by

$$F(x) = \begin{cases} [-1,1] & \text{if } x = 0 \\ \{0\} & \text{if } x \neq 0. \end{cases}$$

Then F is upper semicontinuous at zero but not lower semicontinuous at zero.

(c) Let $X = \mathbb{R}$ be a metric space with the usual metric. The set-valued map $F : X \rightrightarrows X$ defined by

$$F(x) = \begin{cases} \{0\} & \text{if } x = 0 \\ [-1,1] & \text{if } x \neq 0 \end{cases}$$

is lower semicontinuous at zero but not upper semicontinuous at zero.

(d) Let $X = [0,1]$ be a metric space with usual metric. The set-valued map $F : X \rightrightarrows X$ defined by

$$F(x) = \begin{cases} [0,x] & \text{if } 0 \leq x < 1 \\ \{0\} & \text{if } x = 1 \end{cases}$$

is lower semicontinuous but not upper semicontinuous.

Definition 8.8. Let X and Y be metric spaces. A set-valued map $F : X \rightrightarrows Y$ is said to be *continuous* at $x_0 \in X$ if it is both lower semicontinuous as well as upper upper semicontinuous at x_0. It is said to be *continuous* on X if it is continuous at every point $x \in X$.

It is not easy for the readers to prove the set-valued maps are upper semicontinuous or lower semicontinuous by using the definitions of these continuities. So, we have the following characterizations for these definitions.

Theorem 8.10. *Let X and Y be metric spaces. A set-valued map $F : X \rightrightarrows Y$ is lower semicontinuous if and only if $F^{-1}(G)$ is open for every open subset G of Y.*

Proof. Suppose that F is lower semicontinuous and G is an open subset of Y. Clearly $F^{-1}(G)$ is open if it is empty. So, we assume that $F^{-1}(G) \neq \emptyset$. If $x_0 \in F^{-1}(G)$ then $F(x_0) \cap G \neq \emptyset$. Let $y_0 \in F(x_0) \cap G$. Since G is open, it is a neighborhood of each of its points. Therefore, G is a neighborhood of y_0. So, for any $y_0 \in F(x_0)$ and a neighborhood G of y_0, there exists $\delta > 0$ such that $F(x) \cap G \neq \emptyset$ for all $x \in S_\delta(x_0)$. Hence $S_\delta(x_0) \subseteq F^{-1}(G)$. This implies that $F^{-1}(G)$ is open.

Conversely, assume that $F^{-1}(G)$ is open for each open subset G of Y such that $F(x_0) \cap G \neq \emptyset$. Then $x_0 \in F^{-1}(G)$. Since $F^{-1}(G)$ is open, it is a neighborhood of each of its points. We have $F^{-1}(G)$ is a neighborhood of x_0. Then for any $x_0 \in F^{-1}(G)$, there exists $\delta > 0$ such that $S_\delta(x_0) \subseteq F^{-1}(G)$. Therefore, for all $x \in S_\delta(x_0)$ we have $x \in F^{-1}(G)$. This implies that $F(x) \cap G \neq \emptyset$ and hence F is lower semicontinuous.

Corollary 8.1. *Let X and Y be metric spaces. A set-valued map F from X to Y is lower semicontinuous if and only if $F_+^{-1}(H)$ is closed for every closed subset H of Y.*

Proof. It follows from Theorem 8.10 and Theorem 8.6 (a).

Theorem 8.11. *Let X and Y be metric spaces and $F : X \rightrightarrows Y$ be a set-valued map such that $F(x)$ is compact for each $x \in X$. Then F is upper semicontinuous if and only if for each open subset G of Y the set*

$$F_+^{-1}(G) = \{x \in X : F(x) \subseteq G\}$$

is open.

Proof. Suppose that F is upper semicontinuous. If $F_+^{-1}(G) = \emptyset$, then clearly it is open. Assume that $F_+^{-1}(G) \neq \emptyset$ and let $x_0 \in F_+^{-1}(G)$. Then $F(x_0) \subseteq G$. Since G is open, it is a neighborhood of each of its points. Therefore, G is a neighborhood of $F(x_0)$. Then by upper semicontinuity of F, there exists $\delta > 0$ such that $F(x) \subseteq G$ for all $x \in S_\delta(x_0)$. This implies that $S_\delta(x_0) \subseteq F_+^{-1}(G)$ and so $F_+^{-1}(G)$ is an open neighborhood of each of its points and hence $F_+^{-1}(G)$ is open.

Conversely, assume that for each open set G in Y, the set $F_+^{-1}(G)$ is open and that $F(x)$ is compact for each $x \in X$. Let $x_0 \in X$ and let G be an open set containing $F(x_0)$. Since $F(x_0)$ is compact, G is a neighborhood of $F(x_0)$. Then $F_+^{-1}(G)$ is a neighborhood of x_0, and therefore, there exists $\delta > 0$ such that $S_\delta(x_0) \subseteq F_+^{-1}(G)$. Thus for all $x \in S_\delta(x_0)$, $x \in F_+^{-1}(G)$. This implies that $F(x) \subseteq G$ for all $x \in S_\delta(x_0)$ and hence F is upper semicontinuous.

Corollary 8.2. *Let X and Y be metric spaces and $F : X \rightrightarrows Y$ be a set-valued map such that $F(x)$ is compact for each $x \in X$. Then F is upper semicontinuous if and only if for each closed subset H of Y the set $F^{-1}(H)$ is closed.*

Proof. It follows from Theorem 8.11 and Theorem 8.6 (i).

Theorem 8.12. *Let X and Y be metric spaces and $F : X \rightrightarrows Y$ be an upper semicontinuous set-valued map such that for all $x \in X$, $F(x)$ is compact. Then the image $F(K)$ of a compact subset K of X is compact.*

Proof. Let $\{G_\alpha\}_{\alpha \in \Lambda}$ be an open cover of $F(K)$. Since for every $x \in K$ the set $F(x)$ is compact, $F(x)$ can be covered by a finite number of G_α's. Let G_x denote the union of the sets in such a finite family. Then $\{F_+^{-1}(G_x) : x \in K\}$ is an open covering of K and so it contains a finite covering $F_+^{-1}(G_{x_1}), F_+^{-1}(G_{x_2}), \ldots, F_+^{-1}(G_{x_n})$. The sets $G_{x_1}, G_{x_2}, \ldots, G_{x_n}$ cover $F(K)$ and so $F(K)$ can be covered by a finite number of G_α's.

Now, we mention some properties of lower semicontinuous and upper semicontinuous set-valued maps which can be easily proved (proofs are given in [13] and [46]).

1. If F_1 is a lower semicontinuous (upper semicontinuous) set-valued map from a metric space X to another metric space Y and F_2 is another lower semicontinuous (upper semicontinuous) set-valued map from Y to a metric space Z, then the composition product $F = F_2 \circ F_1$ is also lower semicontinuous (upper semicontinuous) set-valued map from X to Z.
2. The arbitrary union $F = \bigcup_{\alpha\in\Lambda} F_\alpha$ of a family of lower semicontinuous set-valued maps F_α from a metric space X to another metric space Y is also lower semicontinuous.
   But the only finite union $F = \bigcup_{i=1}^{n} F_i$ of upper semicontinuous set-valued maps F_i from a metric space X to another metric space Y is upper semicontinuous.
3. The intersection $F = \bigcap_{\alpha\in\Lambda} F_\alpha$ of a family of upper semicontinuous set-valued maps F_α from a metric space X to another metric space Y is upper semicontinuous.
4. The Cartesian product $F = \prod_{i=1}^{n} F_i$ of finite number of lower semicontinuous (upper semicontinuous) set-valued maps F_i from a metric space X to another metric space Y_i is lower semicontinuous (upper semicontinuous) set-valued map from X to $Y = \prod_{i=1}^{n} Y_i$.

Example 8.7. Let $X = [0,\pi]$ and $Y = \mathbb{R}^2$ be the metric spaces with the usual metrics. Consider two set-valued maps $F_1, F_2 : [0,\pi] \rightrightarrows \mathbb{R}^2$ defined by

$$F_1(t) = \{(x,y) \in \mathbb{R}^2 : y \geq 0 \text{ and } x^2 + y^2 \leq 1\} \quad \text{for all } t \in [0,\pi]$$

and

$$F_2(t) = \{(x,y) \in \mathbb{R}^2 : x = \lambda \cos t,\ y = \lambda \sin t,\ \lambda \in [-1,1]\} \quad \text{for all } t \in [0,\pi].$$

F_1 is a constant set-valued map and hence even continuous, F_2 is lower semicontinuous but $F_1 \cap F_2$ is no longer lower semicontinuous (to see it consider $t = 0$ or $t = \pi$).

Definition 8.9. Let X and Y be metric spaces. A set-valued map F from X to Y is said to be *closed* if whenever $x_0 \in X$, $y_0 \in Y$, $y_0 \notin F(x_0)$, there exist two neighborhoods $U(x_0)$ of x_0 and $V(y_0)$ of y_0 such that $F(x) \cap V(y_0) = \emptyset$ for all $x \in U(x_0)$.

Example 8.8. (a) If $f : X \times Y \to \mathbb{R}$ is a continuous single-valued function from a metric space X to another metric space Y, then the set-valued map F from X to Y defined by $F(x) = \{y \in Y : f(x,y) \leq 0\}$ is closed.

(b) If f is a continuous function from a metric space (X,d) to $\mathbb{R}$, then the set-valued map F from a metric space X to another metric space Y defined by

$$F(x) = \{y \in Y : d(x,y) - f(x) \leq 0\}$$

is closed.

We state some useful characterizations of closed set-valued maps.

1. F is closed if and only if its graph $Graph(F)$ is a closed set.
2. F is closed if and only if for any sequences $\{x_n\}$ and $\{y_n\}$ such that $x_n \to x_0$, $y_n \to y_0$ and $y_n \in F(x_n)$ for all n, we have $y_0 \in F(x_0)$.

Example 8.9. Let $X = \mathbb{R}_+$ be a metric space with the usual metric. Consider the set-valued map $F : X \rightrightarrows X$ defined by

$$F(x) = \left(\frac{1}{x}, +\infty\right) \quad \text{for all } x > 0.$$

Then the graph of F is the set

$$Graph(F) = \{(x,y) \in X \times X : xy \geq 1\}.$$

The set-valued map F is closed as the $Graph(F)$ is closed.

Theorem 8.13. *Let X and Y be metric spaces and $F : X \rightrightarrows Y$ be a closed set-valued map. If K is a compact subset of X, then the set $F(K)$ is closed.*

Proof. Let $y_n \in F(K)$ for all $n = 1,2,\ldots$, and $y_n \to y$. Suppose that x_n is an element of $F^{-1}(y_n) \cap K$. Without loss of generality, we may assume that $\{x_n\}$ converges to some point x. It follows from the definition of closedness of F that $y \in F(x)$ and hence $y \in F(K)$.

Corollary 8.3. *Let X and Y be metric spaces and $F : X \rightrightarrows Y$ be a closed set-valued map. Then the set $F(x)$ is closed for all $x \in X$.*

The converse of above corollary is not valid as shown by the following example.

Example 8.10. Let $X = \mathbb{R}_+$ be a metric space with the usual metric and $F : X \rightrightarrows X$ be a set-valued map defined by

$$F(x) = \begin{cases} [0,1] & \text{when } x \neq 0 \\ \{0\} & \text{when } x = 0. \end{cases}$$

The set $F(x)$ is closed for all $x \in X$. However, the set-valued map F is not closed.

Remark 8.2. If F is a closed set-valued map, then the image of a closed (respectively, compact) set need not be closed (respectively, compact) (even if the set $F(x)$ is compact for all $x \in X$).

Example 8.11. Let $X = [1, +\infty)$ and $Y = \mathbb{R}_+$ be metric spaces with the usual metrics and $F : X \rightrightarrows Y$ be a set-valued map defined as $F(x) = \left[\frac{1}{x}, 1\right]$ for all $x \in X$. Then F is closed but the image $F(X)$ of the closed set X is not closed.

Example 8.12. Let $X = \mathbb{R}_+$ be a metric space with the usual metric and $F : X \rightrightarrows Y$ be a set-valued map defined as

$$F(x) = \begin{cases} \{0\} & \text{when } x = 0 \\ \left[-\frac{1}{x}, \frac{1}{x} + 1\right] & \text{when } x \neq 0. \end{cases}$$

Then the map F is closed but the image of the compact set $[0,1]$ is not compact. Note that the set $F(x)$ is compact for all $x \in X$.

The composition of two closed set-valued maps need not be closed.

Example 8.13. Let $X = Y = Z = \mathbb{R}_+$ be metric spaces with the usual metric. Consider the set-valued maps $F_1 : X \rightrightarrows Y$ and $F_2 : Y \rightrightarrows Z$ defined by

$$F_1(x) = \begin{cases} \{\frac{1}{x}\} & \text{if } x \neq 0 \\ \{0\} & \text{if } x = 0 \end{cases}$$

and

$$F_2(x) = \begin{cases} \{\frac{1}{x}\} & \text{if } x \neq 0 \\ \{1\} & \text{if } x = 0, \end{cases}$$

respectively. Then F_1 and F_2 are closed set-valued maps. However, the set-valued map $F_2 \circ F_1$ given by

$$(F_2 \circ F_1)(x) = \begin{cases} \{x\} & \text{if } x \neq 0 \\ \{1\} & \text{if } x = 0, \end{cases}$$

is not closed.

Definition 8.10. Let X and Y be metric spaces. A set-valued map $F : X \rightrightarrows Y$ is said to be *bounded* if the image of a bounded set is bounded.

Theorem 8.14. *Let X, Y and Z be metric spaces, and $F_1 : X \rightrightarrows Y$ and $F_2 : Y \rightrightarrows Z$ be closed set-valued maps. If F_1 is bounded, then $F_2 \circ F_1$ is a closed set-valued map from X to Z.*

Proof. Consider the sequences $\{x_n\}$ in X and $\{z_n\}$ in $(F_2 \circ F_1)(x_n)$ such that $x_n \to x$ and $z_n \to z$. Then we have to show that $z \in (F_2 \circ F_1)(x)$.

For every n, there exists an element $y_n \in F_1(x_n)$ such that $z_n \in F_2(y_n)$. Since every convergent sequence is bounded, $\{x_n\}$ is bounded and the boundedness of F_1 implies that the sequence $\{y_n\}$ is bounded. Hence, without loss of generality, we may assume that it converges to some point y. The closedness of F_1 and F_2 implies that $y \in F_1(x)$ and $z \in F_2(y)$, that is, $z \in (F_2 \circ F_1)(x)$.

Theorem 8.15. *Let X and Y be metric spaces. If $F : X \rightrightarrows Y$ is an upper semicontinuous set-valued map with compact values, then it is closed.*

Proof. Let $x_0 \in X$ be an arbitrary and assume that $y_0 \notin F(x_0)$. Since $F(x_0)$ is compact, there exist an open set G in Y containing $F(x_0)$ and a neighborhood $V(y_0)$ of y_0 such that $G \cap V(y_0) = \emptyset$. Since F is upper semicontinuous, there exists a neighborhood $U(x_0)$ of x_0 such that $F(x) \subseteq G$ for all $x \in U(x_0)$. Then $F(x) \cap V(y_0) = \emptyset$ for all $x \in U(x_0)$. Therefore, F is closed.

Remark 8.3. In general if $f : X \to Y$ is a single-valued continuous map from X onto Y, then the inverse map $f^{-1} : Y \rightrightarrows X$ is a set-valued map and it has a closed graph. But it is not necessarily upper semicontinuous.

Theorem 8.16. *Let X and Y be metric spaces and $F : X \rightrightarrows Y$ be a closed set-valued map. If Y is compact, then F is upper semicontinuous.*

Proof. Assume to the contrary that F is not upper semicontinuous. Then there exists an open neighborhood $V_{F(x)}$ of $F(x)$ in Y such that for every open neighborhood U_x of x in X we have $F(U_x) \not\subset V_{F(x)}$.

We take $U_x = S_{1/n}(x)$, $n = 1, 2, \ldots$ Then for each n, we get a point $x_n \in S_{1/n}(x)$ such that $F(x_n) \not\subset V_{F(x)}$. For every n, let $y_n \in F(x_n)$ such that $y_n \notin V_{F(x_n)}$. Then we

have $x_n \to x \in X$ and $\{y_n\} \subset Y$. Since Y is compact, we may assume without loss of generality that $y_n \to y \in Y$. We see that $y \notin V_{F(x)}$. Therefor for each n, we have $(x_n, y_n) \in Graph(F)$ and $\{(x_n, y_n)\} \to (x, y)$. So $(x, y) \in Graph(F)$ because $Graph(F)$ is a closed subset of $X \times Y$. So, $y \in F(x)$ and hence $y \in V_{F(x)}$. This contradicts to the fact that $y \notin V_{F(x)}$ and the proof is complete.

If Y is not compact, then the conclusion of Theorem 8.16 does not hold.

Example 8.14. Let $X = \mathbb{R}$ be a metric space with the usual metric and $F : X \rightrightarrows X$ be a set-valued map defined as

$$F(x) = \begin{cases} \{\frac{1}{x}\} & \text{when } x \neq 0 \\ \{0\} & \text{when } x = 0. \end{cases}$$

Then F is closed but not upper semicontinuous.

8.3 Nadler's Fixed Point Theorem

Let X be a metric space. We use the following notions:

2^X = Set of all subsets of X;

2^X_{cl} = Set of all nonempty closed and bounded subsets of X;

2^X_q = Set of all nonempty compact subsets of X.

Evidently, we have $2^X_q \subseteq 2^X_{cl}$.

Throughout this section, unless otherwise specified, we assume that X is a metric space with the metric d. For any two nonempty elements M and N of 2^X, we define the number $d(M,N)$ as follows: let $x \in M$ and set

$$d(x,N) = \inf_{y \in N} d(x,y)$$

and

$$d(M,N) = \sup_{x \in M} d(x,N) = \sup_{x \in M} \inf_{y \in N} d(x,y).$$

By convention, $\inf \emptyset = \infty$.

The number $d(M,N)$ has the following properties.

Proposition 8.1. *Let M, N, Q be nonempty bounded subsets of X. Then*

(a) $d(M,N) = 0$ if and only if $M \subseteq \overline{N}$;
(b) $d(M,N) \leq d(M,Q) + d(Q,N)$.

Proof. The first assertion follows obviously from the definition of the number $d(M,N)$ and from the properties of the closure of a bounded set in a metric space.

The second assertion is proved as follows:

From the triangle inequality, we have

$$d(m,n) \leq d(m,q) + d(q,n) \quad \text{for all } m \in M,\ n \in N,\ q \in Q.$$

This implies that

$$\inf_{n\in N} d(m,n) \le d(m,q) + \inf_{n\in N} d(q,n)$$

holds for all $q \in Q$. This gives further that

$$\inf_{n\in N} d(m,n) = d(m,N) \le d(m,q) + d(q,N)$$

and since this is valid for all $q \in Q$, we get

$$d(m,N) \le d(m,Q) + d(Q,N).$$

By taking $\sup_{m\in M}$, we obtain

$$d(M,N) \le d(M,Q) + d(Q,N).$$

The proof is complete.

The above definition of d does not define a metric. There are various problems. For example, in $\mathbb{R}$, what is the distance between $\{0\}$ and $[0,\infty)$? It is infinite. That is not allowed in the definition of a metric. Therefore, we will restrict the use of d to bounded sets. What is the distance $d(\emptyset,\{0\})$? Again, infinite. So we will restrict the use of d to nonempty sets. What is the distance $d((0,1),[0.1])$? Now the distance is zero, even though the two sets are not equal. Therefore, we will restrict the use of d to closed sets. In general,

$$d(M,N) \ne d(N,M).$$

To avoid the difficulty posed by the above remarks, we define *Hausdorff metric* $\mathscr{H}$ on the family of all nonempty closed and bounded subsets of a metric space (X,d) by

$$\mathscr{H}(M,N) = \max\{d(M,N), d(N,M)\}. \tag{8.1}$$

$\mathscr{H}(M,N)$ is the distance from the set M to the set N and it is called *Hausdorff distance* from set M to a set N.

Proposition 8.2. *Let (X,d) be a metric space, $M,N \in 2^X_q$ and $r > 0$ a given number. Then $M \subseteq S_r(N)$ and $N \subseteq S_r(M)$ if and only if $\mathscr{H}(M,N) \le r$.*

Proof. We first show that $M \subseteq S_r(N)$ if and only if $d(M,N) \le r$. Assume that $d(M,N) \le r$. Since $d(M,N) = \sup_{x\in M} d(x,N) \le r$, we obtain $d(x,N) \le r$ for all $x \in M$. Since N is compact, there is a point $y \in N$ such that $d(x,y) \le r$. Hence for each $x \in M$ belongs to a closed sphere of radius r with center at $y \in N$, that is, $x \in S_r(N)$ and hence $M \subseteq S_r(N)$.

Suppose that $M \subseteq S_r(N)$. Then for all $x \in M$, there is a point $y \in N$ such that $d(x,y) \le r$. Thus for all $x \in M$, we have $d(x,N) \le d(x,y) \le r$. Since $d(M,N) = \sup_{x\in M} d(x,N)$, we conclude that $d(M,N) \le r$.

By interchanging the sets M and N, we deduce that $N \subseteq S_r(M)$ if and only if $d(N,M) \le r$.

Now $\mathscr{H}(M,N) = \max\{d(M,N), d(N,M)\} \le r$ if and only if $d(M,N) \le r$ and $d(N,M) \le r$ if and only if $M \subseteq S_r(N)$ and $N \subseteq S_r(M)$.

In view of above proposition, (8.1) can be rewritten as

$$\mathscr{H}(M,N) = \inf\{r > 0 : M \subseteq S_r(N) \text{ and } N \subseteq S_r(M)\}$$

for all $M,N \in 2^X_q$.

Theorem 8.17. *Let (X,d) be a metric space. Then $\mathscr{H}(.,.)$ is a metric on 2^X_{cl}.*

Proof. Let $M,N \in 2^X_{cl}$. Since M and N are closed and bounded, we have $\mathscr{H}(M,N) = d(x,y)$ for some $x \in M$ and $y \in N$. Hence $0 \leq \mathscr{H}(M,N) < \infty$. Clearly, $\mathscr{H}(M,M) = \max\{d(M,M),d(M,M)\} = d(M,M) = \sup_{x\in M} d(x,M) = 0$.

Conversely, suppose that $\mathscr{H}(M,N) = 0$. If $x \in M$ then $d(x,N) = 0$. Since N is closed, by Theorem 1.5, $x \in N$. Therefore, $M \subseteq N$. Similarly $N \subseteq M$, so $M = N$.

Clearly, $\mathscr{H}(M,N) = \mathscr{H}(N,M)$.

Finally, we prove the triangle inequality

$$\mathscr{H}(M,N) \leq \mathscr{H}(M,Q) + \mathscr{H}(Q,N) \quad \text{for all } M,N,Q \in 2^X_{cl}.$$

From Proposition 8.1 (b) we have

$$d(M,N) \leq d(M,Q) + d(Q,N) \quad \text{and} \quad d(N,M) \leq d(N,Q) + d(Q,M),$$

for all $M,N,Q \in 2^X_{cl}$. Hence,

$$\begin{aligned}
\mathscr{H}(M,N) &= \max\{d(M,N),d(N,M)\} \\
&\leq \max\{d(M,Q) + d(Q,N), d(N,Q) + d(Q,M)\} \\
&\leq \max\{d(M,Q),d(Q,M)\} + \max\{d(Q,N),d(N,Q)\} \\
&= \mathscr{H}(M,Q) + \mathscr{H}(Q,N).
\end{aligned}$$

This completes the proof.

Theorem 8.18. *For any metric space (X,d), the following assertions hold:*

(a) If (X,d) is a complete metric space, then $\left(2^X_q, \mathscr{H}\right)$ is a complete metric space.
(b) If (X,d) is a separable metric space, then $\left(2^X_q, \mathscr{H}\right)$ is a separable metric space.

For the proof of above theorem, see [51, pp. 297].

Note that the topology on 2^X_q derived from the Hausdorff distance $\mathscr{H}$ is not determined by the metric topology of (X,d). Two topologically equivalent metrics d and d' may lead to very different topologies on 2^X_q by the Hausdorff distance procedure. It follows from the example given below.

Example 8.15. Let $X = \mathbb{R}_+ = [0,+\infty)$,

$$d(x,y) = \left|\frac{x}{(1+x)} - \frac{y}{(1+y)}\right|$$

and

$$d'(x,y) = \min\{1,|x-y|\}.$$

The metrics d and d' define the same topology on $\mathbb{R}_+$ but the topologies of the Hausdorff distance on 2^X_q are different, that is, the set $\mathbb{N}$ of natural numbers belongs to the closure of the set of all finite subsets of $\mathbb{N}$ in the first space but not in the second.

Above example shows that 2^X_q is not a closed subset of $\left(2^X_{cl}, \mathscr{H}\right)$ in general. However the following result holds.

Proposition 8.3. *If (X,d) is a complete metric space, then 2^X_q is a closed subset of $\left(2^X_{cl}, \mathscr{H}\right)$.*

For the proof of above theorem, see [46, pp. 24].

Definition 8.11. Let (X,d_1) and (Y,d_2) be metric spaces. A set-valued map $F : X \to 2^Y_q$ is said to be *$\mathscr{H}$-continuous* if for any $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $x,y \in X$

$$\mathscr{H}(F(x),F(y)) < \varepsilon \quad \text{whenever} \quad d_1(x,y) < \delta.$$

Theorem 8.19. *Let X and Y be metric spaces. A mapping $F : X \to 2^Y_q$ is $\mathscr{H}$-continuous if and only if the set-valued map $F : X \rightrightarrows Y$ is compact-valued and upper semicontinuous as well as lower semicontinuous on X.*

Proof. Let F be $\mathscr{H}$-continuous and G be an open subset of Y. We first show that F is upper semicontinuous by proving that the set $F_+^{-1}(G) = \{x \in X : F(x) \subseteq G\}$ is open in X.

Let $x_0 \in F_+^{-1}(G)$. Then $F(x_0) \subseteq G$. Since $F(x_0)$ is compact, there exists $\varepsilon > 0$ such that $S_\varepsilon(F(x_0)) \subseteq G$. Since F is $\mathscr{H}$-continuous, we can find a $\delta > 0$ such that for all $x \in S_\delta(x_0)$, we have

$$\mathscr{H}(F(x_0),F(x)) < \varepsilon.$$

It implies that $F(x) \subseteq S_\varepsilon(F(x_0)) \subseteq G$, and so, $S_\delta(x_0) \subseteq F_+^{-1}(G)$. Thus $F_+^{-1}(G)$ is open in X.

Now we show that F is lower semicontinuous by proving that the set $F^{-1}(G) = \{x \in X : F(x) \cap G \neq \emptyset\}$ is open in X.

Let $x_0 \in F^{-1}(G)$. Then $F(x_0) \cap G \neq \emptyset$ and so we let $y_0 \in F(x_0) \cap G$. Since G is open and $y_0 \in G$, there exists $\varepsilon > 0$ such that $S_\varepsilon(y_0) \subseteq G$. Since F is $\mathscr{H}$-continuous, we can find a $\delta > 0$ such that for all $x \in S_\delta(x_0)$, we have

$$\mathscr{H}(F(x_0),F(x)) < \frac{\varepsilon}{2}.$$

We claim that $F(x) \cap S_\varepsilon(y_0) \neq \emptyset$. Assume to the contrary that $F(x) \cap S_\varepsilon(y_0) = \emptyset$. On the other hand $F(x_0) \subset S_{\varepsilon/2}(F(x))$. Therefore, $y_0 \in S_{\varepsilon/2}(F(x))$ and there exists $z_0 \in F(x)$ such that $d(y_0,z_0) < \varepsilon/2$. It implies that $z_0 \in S_\varepsilon(y_0)$ and we obtain a contradiction.

Conversely, assume that F is both upper semicontinuous as well as lower semicontinuous. Let $\varepsilon > 0$ and $x_0 \in X$. Let $U = S_\varepsilon(F(x_0))$. Then the sets $F^{-1}(U)$ and $F_+^{-1}(U)$ are open and $x_0 \in F^{-1}(U) \cap F_+^{-1}(U)$. Let $V = F^{-1}(U) \cap F_+^{-1}(U)$. Then V is an open neighborhood of x_0 such that $F(x) \subset S_\varepsilon(F(x_0))$ for all $x \in V$. We are looking for $\delta > 0$ such that $S_\delta(x_0) \subset V$ and $F(x_0) \subset S_\varepsilon(F(x))$ for every $x \in S_\delta(x_0)$.

To find such δ we cover the compact set by n open spheres $S_\varepsilon(y_i)$, $i = 1,2,\ldots,n$. Then $F(x_0) = \bigcup_{i=1}^n S_\varepsilon(y_i) \subset S_{\varepsilon/2}(F(x_0))$. Since F is lower semicontinuous, there are open spheres $S_{\delta_i}(x_0) \subset V$ such that

$$F(x) \cap S_{\varepsilon/2}(y_i) \neq \emptyset \quad \text{for all } x \in S_{\delta_i}(x_0).$$

Let $\delta = \min\{\delta_1, \delta_2, \ldots, \delta_n\}$. Then $S_\delta(x_0) \subset V$ and any $y \in F(x_0)$ belongs to $S_{\varepsilon/2}(y_i)$ for some i. Furthermore, we know that for any $x \in S_\delta(x_0)$, $F(x) \cap S_{\varepsilon/2}(y_i) \neq \emptyset$ for all $i = 1, 2, \ldots, n$. Thus for every $x \in S_\delta(x_0)$ and $y \in F(x_0)$, there exists $i = 1, 2, \ldots, n$ such that

$$\rho(y, F(x)) \leq d(y, y_i) + \rho(y_i, F(x_i)) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Therefore for all $x \in S_\delta(x_0)$, we obtain

$$F(x_0) \subset S_\varepsilon(F(x)).$$

This completes the proof.

Note that if F is only upper semicontinuous or lower semicontinuous, then F is not $\mathscr{H}$-continuous in general.

Example 8.16. Let $X = [0,1]$ be a metric space with the usual metric. Consider the set-valued map $F : X \rightrightarrows X$ defined by

$$F(x) = \begin{cases} \{1\} & \text{if } x < \frac{1}{2} \\ [0,1] & \text{if } x = \frac{1}{2} \\ \{0\} & \text{if } x > \frac{1}{2}. \end{cases}$$

Then F is upper semicontinuous but not $\mathscr{H}$-continuous.

The following example shows that the only lower semicontinuity does not imply $\mathscr{H}$-continuity.

Example 8.17. Let $X = Y = [0,1]$ be a metric space with the usual metric and $F : [0,1] \to 2_q^{[0,1]}$ be defined by

$$F(x) = \begin{cases} [0,1] & \text{when } x \neq 0 \\ \{0\} & \text{when } x = 0. \end{cases}$$

Then F is lower semicontinuous with compact values but

$$\mathscr{H}(F(0), F(x)) = 1 \quad \text{for all } x \neq 0,$$

so F is not $\mathscr{H}$-continuous.

Theorem 8.20. *Let (X,d) be a metric space. If $F : X \to 2_{cl}^X$ is $\mathscr{H}$-continuous, then the set-valued map $F : X \rightrightarrows X$ is lower semicontinuous.*

For the proof of above theorem, see [46, pp. 95].

Note that the $\mathscr{H}$-continuity of a set-valued map $F : X \to 2_{cl}^X$ does not imply upper semicontinuity.

Example 8.18. Let $X = \mathbb{R}$ be the Euclidean space of real numbers and let $Y = \mathbb{R}^2$ be equipped with the bounded metric d^* defined as

$$d^*(x,y) = \frac{d(x,y)}{1+d(x,y)},$$

where d is the usual metric on $\mathbb{R}^2$.

Consider the mapping $F : \mathbb{R} \to 2^{\mathbb{R}^2}_{cl}$ defined as

$$F(t) = \{(t,y) : y \in \mathbb{R}\} \quad \text{for all } t \in \mathbb{R}.$$

Then we have

$$\mathscr{H}(F(t),F(s)) \leq 2|t-s| \quad \text{for all } t,s \in \mathbb{R}$$

and so F is $\mathscr{H}$-continuous.

Let $U = \{(x,y) \in \mathbb{R}^2 : |y| < 1/x \text{ or } x = 0\}$. Then U is an open subset of $\mathbb{R}^2$ but $F_+^{-1}(U) = \{0\}$ is not open in $\mathbb{R}$. Consequently, F is not upper semicontinuous.

Definition 8.12. Let F be a set-valued map from a metric space X to itself. A point $x \in X$ is said to be a *fixed point* of F if $x \in F(x)$.

Definition 8.13. Let (X,d_1) and (Y,d_2) be two metric spaces. A map $F : X \to 2^Y_{cl}$ is said to be *set-valued Lipschitz map* if there exists a constant $\alpha \geq 0$ such that

$$\mathscr{H}(F(x),F(y)) \leq \alpha d_1(x,y) \quad \text{for all } x,y \in X.$$

The constant α is called a *Lipschitz constant* for F. If F has Lipschitz constant $\alpha < 1$, then F is called a *set-valued contraction map*. If F has Lipschitz constant $\alpha = 1$, then F is called *nonexpansive*.

Example 8.19. Let $X = [0,1]$ be a metric space with the usual metric and let $f : X \to X$ be given by

$$f(x) = \begin{cases} \frac{1}{2}x + \frac{1}{2} & \text{if } 0 \leq x \leq \frac{1}{2} \\ -\frac{1}{2}x + 1 & \text{if } \frac{1}{2} \leq x \leq 1. \end{cases}$$

Define $F : X \rightrightarrows X$ by $F(x) = \{0\} \cup \{f(x)\}$ for all $x \in X$. Then F is a set-valued contraction map and the fixed points of F are 0 and $\frac{2}{3}$.

Remark 8.4. Let $F : X \to 2^Y_{cl}$ by a set-valued Lipschitz map with Lipschitz constant α and let $G : X \to 2^Y_{cl}$ be another set-valued Lipschitz map with Lipschitz constant β. If $F \cup G : X \to 2^Y_{cl}$ is given by $(F \cup G)(x) = F(x) \cup G(x)$ for all $x \in X$, then $F \cup G$ is a set-valued Lipschitz map with Lipschitz constant $\max\{\alpha,\beta\}$.

But the intersection of two set-valued contraction maps need not be a contraction.

Example 8.20. Let $I^2 = \{(x,y) : 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1\}$, let $F : I^2 \to 2^{I^2}_{cl}$ be defined by

$F(x,y) =$ line segment in I^2 from the point $(\frac{1}{2}x,0)$ to the point $(\frac{1}{2}x,1)$ for each $(x,y) \in I^2$,

and let $G : I^2 \to 2^{I^2}_{cl}$ be defined by

$G(x,y) =$ line segment in I^2 from the point $(\frac{1}{2}x,0)$ to the point $(\frac{1}{3}x,1)$ for each $(x,y) \in I^2$.

F and G are multivalued contraction mappings and $F \cap G$, which is given by

$$(F \cap G)(x,y) = \begin{cases} \{(\frac{1}{2}x,0)\} & \text{if } x \neq 0 \\ \{(x,y) \in I^2 : x = 0\} & \text{if } x = 0, \end{cases}$$

for all $(x,y) \in I^2$, is not a contraction.

In 1969, Nadler [64] established the following analogue of Banach contraction theorem for set-valued maps.

Theorem 8.21 (Nadler's Theorem). *Let (X,d) be a complete metric space. If $F : X \to 2^X_{cl}$ is a set-valued contraction map then F has a fixed point.*

Proof. Let α be a Lipschitz constant for F, and let $x_0 \in X$. Choose $x_1 \in F(x_0)$. Since $F(x_0), F(x_1) \in 2^X_{cl}$ and $x_1 \in F(x_0)$, there is a point $x_2 \in F(x_1)$ such that

$$d(x_1,x_2) \leq \mathscr{H}(F(x_0),F(x_1)) + \alpha.$$

Since $F(x_1), F(x_2) \in 2^X_{cl}$ and $x_2 \in F(x_1)$, there is a point $x_3 \in F(x_2)$ such that

$$d(x_2,x_3) \leq \mathscr{H}(F(x_1),F(x_2)) + \alpha^2.$$

Continuing in this fashion, we produce a sequence $\{x_n\}_{n=1}^{\infty}$ of points of X such that $x_{n+1} \in F(x_n)$ and

$$d(x_n,x_{n+1}) \leq \mathscr{H}(F(x_{n-1}),F(x_n)) + \alpha^n \quad \text{for all } n \geq 1.$$

Note that

$$\begin{aligned} d(x_n,x_{n+1}) &\leq \mathscr{H}(F(x_{n-1}),F(x_n)) + \alpha^n \\ &\leq \alpha d(x_{n-1},x_n) + \alpha^n \\ &\leq \alpha\left[\mathscr{H}(F(x_{n-2}),F(x_{n-1})) + \alpha^{n-1}\right] + \alpha^n \\ &\leq \alpha^2 d(x_{n-2},x_{n-1}) + 2\alpha^n \\ &\;\;\vdots \\ &\leq \alpha^n d(x_0,x_1) + n\alpha^n \end{aligned}$$

for all $n \geq 1$. This implies that for any n and m,

$$\begin{aligned} d(x_n,x_{n+m}) &\leq d(x_n,x_{n+1}) + d(x_{n+1},x_{n+2}) + \cdots + d(x_{n+m-1},x_{n+m}) \\ &\leq \alpha^n d(x_0,x_1) + n\alpha^n + \alpha^{n+1} d(x_0,x_1) + (n+1)\alpha^{n+1} + \\ &\quad \cdots + \alpha^{n+m-1} d(x_0,x_1) + (n+m-1)\alpha^{n+m-1} \\ &= \sum_{k=n}^{n+m-1} \alpha^k d(x_0,x_1) + \sum_{k=n}^{n+m-1} k\alpha^k. \end{aligned}$$

It follows that the sequence $\{x_n\}$ is a Cauchy sequence. Since (X,d) is complete, the sequence $\{x_n\}$ converges to some point $x \in X$. Since F is $\mathscr{H}$-continuous,

$$\lim_{n\to\infty} \mathscr{H}(F(x_n),F(x)) = 0.$$

Also, since $x_n \in F(x_{n-1})$, we have

$$\lim_{n\to\infty} \rho(x_n, F(x)) = \lim_{n\to\infty} \inf\{d(x_n, y) : y \in F(x)\} = 0.$$

This implies that

$$\rho(x, F(x)) = \inf\{d(x, y) : y \in F(x)\} = 0,$$

and since $T(x)$ is closed, by Theorem 1.5, $x \in T(x)$.

8.4 Some Fixed Point Theorems for Set-Valued Maps

The following theorem is established by Dancs, Hegedüs and Medvegyev [30]. We call it DHM theorem.

Theorem 8.22 (DHM Theorem). *Let (X, d) be a complete metric space and $F : X \rightrightarrows X$ be a set-valued map such that the following conditions hold.*

(i) For all $x \in X$, $F(x)$ is closed;
(ii) For all $x \in X$, $x \in F(x)$;
(iii) For all $x, y \in X$, $y \in F(x)$ implies $F(y) \subseteq F(x)$;
(iv) For any sequence $\{x_n\}$ in X such that

$$x_2 \in F(x_1), x_3 \in F(x_2), \ldots, x_n \in F(x_{n-1}), \ldots, \tag{8.2}$$

the distance $d(x_n, x_{n+1})$ tends to zero as $n \to +\infty$.

Then the set-valued map F has a fixed point $\bar{x} \in X$ in the sense that $F(\bar{x}) = \{\bar{x}\}$.

The iterate (8.2) is called the *generalized Picard-iteration.*

Proof. Since d satisfies condition (iv), the equivalent metric $d' = \dfrac{d}{(1+d)}$ also satisfies the condition (iv). So we can suppose that d is bounded on X. Let $\delta(A)$ be the diameter of a subset A of X. Because of condition (ii), $F(x) \neq \emptyset$ for all $x \in X$, and we can construct a generalized Picard-iteration such that $x_1 = \bar{x}$, $x_n \in F(x_{n-1})$ and

$$d(x_n, x_{n-1}) \geq \frac{\delta(F(x_{n-1}))}{2} - \frac{1}{2^{n-1}}.$$

Hence from conditions (iii) and (iv), we have $F(x_{n-1}) \supseteq F(x_n)$ and the diameter of $F(x_n)$, $\delta(F(x_n)) \to 0$ as $n \to +\infty$. Since X is complete, $\{F(x_n)\}$ is a decreasing sequence of nonempty closed sets such that $\delta(F(x_n)) \to 0$ as $n \to +\infty$, by Cantor's intersection theorem

$$\bigcap_{n=1}^{\infty} F(x_n) = \{\bar{x}\}.$$

The point $\bar{x}$ is a fixed point, since on the other hand $\bar{x} \in F(x_n)$ and (iii) imply $F(\bar{x}) \subseteq \bigcap_{n=1}^{\infty} F(x_n) = \{\bar{x}\}$. From condition (ii), we have $\{\bar{x}\} \subseteq F(\bar{x})$ and hence $F(\bar{x}) = \bar{x}$.

Theorem 8.23. *If the metric space (X,d) is noncomplete, then there is a set-valued map $F : X \rightrightarrows X$ which satisfies conditions (i)–(iv) of Theorem 8.22 but has no fixed point.*

Proof. From the assumption, there is a sequence $X = N_0 \supseteq N_1 \supseteq \cdots \supseteq N_n \supseteq \cdots$ of nonempty closed sets in X such that $\delta(N_n) \to 0$, but $\bigcap_{n=1}^{\infty} N_n = \emptyset$. Define the set-valued map $F : X \rightrightarrows X$ by

$$F(x) = N_{n+1} \cup \{x\} \quad \text{if } x \in N_n \text{ and } x \notin N_{n+1}.$$

The set-valued map F satisfies the assumptions of Theorem 8.22, but has no fixed point, since if $\bar{x}$ were a fixed point of F, we would have $F(\bar{x}) = \{\bar{x}\} = N_{n_0} \cup \{\bar{x}\}$ implying $N_{n_0} = \emptyset$. This contradicts to our assumption.

By using Caristi's fixed point theorem, we obtain the following fixed point theorem for set-valued maps.

Theorem 8.24 (Caristi and Kirk Fixed Point Theorem). *Let (X,d) be a complete metric space and let $f : X \to (-\infty, +\infty]$ be a proper, bounded below and lower semi-continuous functional on X. Let $T : X \rightrightarrows X$ be a set-valued map such that for each $x \in X$, there exists $y \in T(x)$ satisfying*

$$f(y) + d(x,y) \leq f(x).$$

Then there exists a fixed point $\bar{x} \in X$ of T, that is, $\bar{x} \in T(\bar{x})$ and $f(\bar{x}) < +\infty$.

Proof. For each $x \in X$, let $\varphi(x) = y$, where y is an element in X such that $y \in T(x)$ and $f(y) + d(x,y) \leq f(x)$. Then φ is a mapping from X into itself and satisfying

$$f(\varphi(x)) + d(x, \varphi(x)) \leq f(x) \quad \text{for all } x \in X.$$

By Caristi's fixed point theorem, there exists $\bar{x} \in X$ such that $\bar{x} = \varphi(\bar{x}) \in T(\bar{x})$ and $f(\bar{x}) < +\infty$.

Chapter 9
Ekeland's Variational Principle and its Applications

The variational principle which provides an approximate minimizer of a bounded below and lower semicontinuous function in a given neighborhood of a point, is discovered by Ekeland [33] in 1972 (see also, [34, 35]). It is known as Ekeland's variational principle (in short, EVP). The localization property in EVP is very useful and explains the importance of this result. EVP is among the most important results obtained in nonlinear analysis. It is appeared as one of the most useful tools to solve the problems from optimization, optimal control theory, game theory, nonlinear equations, dynamical systems, etc; See for example [4, 6, 7, 33, 34, 35, 36, 39, 45, 69, 72, 87] and the references therein. Since the discovery of EVP, there have also appeared many extensions or equivalent formulations of EVP; See for example [1, 7, 16, 42, 48, 49, 50, 54, 63, 69, 70, 71, 72, 83, 84, 85] and the references therein. In 1981, Sullivan [83] established the validity of EVP statement on a metric space (X,d) is equivalent to the completeness of the metric space (X,d). In 1982, McLinden [63] showed that how EVP, or more precisely the augmented form of it provided by Rockafellor [74], can be adapted to extremum problems of minimax type. Many famous results, namely, the Caristi-Kirk fixed point theorem [22, 23], the Petal theorem [72] and the Daneš drop theorem [31], were discovered around the same time but independently of each other. Soon later, it is found that all these results are equivalent to EVP. In 1986, Penot [72] proved that EVP is equivalent to some most famous results in the nonlinear analysis, namely, Caristi's fixed point theorem [22] and petal theorem. In the same paper, some areas of applications of a geometric result of Daneš [31] known as drop theorem are mentioned. For further detail on the equivalence of these results, we refer to [1, 32, 42, 48, 70, 82] and the references therein.

In this chapter, we present several forms of Ekeland's variational principle, Takahashi's minimization theorem and some applications to fixed point theory. We discuss equilibrium problem which is a unified model of several problems, namely, minimization problem, saddle point problem, Nash equilibrium problem, fixed point problem, variational inequality problem, etc. We present equilibrium version of EVP and prove the equivalence with some other problems. We use such EVP to establish the existence of a solution of equilibrium problem without convexity assumption on the underlying set.

9.1 Ekeland Variational's Principle in Complete Metric Spaces

Let X be a nonempty set. A function $f : X \to \mathbb{R} \cup \{+\infty\}$ is said to be *proper* if $f(x) \neq +\infty$ for all $x \in X$. The domain of f, denoted by $dom(f)$, is defined by

$$dom(f) = \{x \in X : f(x) < +\infty\}.$$

Let (X,d) be a metric space, K a nonempty subset of X, and $f: X \to \mathbb{R} \cup \{+\infty\}$ a proper functional. It is well known that if K is compact and f is lower semicontinuous, then the following constrained optimization problem (in short, COP)

$$\inf_{x \in K} f(x)$$

has a solution. Not only this, the solution set of COP is compact. Now the question is "Can we achieve the infimum of the following optimization problem (in short, OP)

$$\inf_{x \in X} f(x)$$

or of the COP without compactness assumption?" The answer is "yes". But we need some kind of coercivity assumption as well as convexity structure on K.

But we can always obtain an *approximately ε-solution*, that is, a point x_ε for $\varepsilon > 0$ satisfying

$$\inf_X f \le f(x_\varepsilon) \le \inf_X f + \varepsilon,$$

where $\inf_X f := \inf_{x \in X} f(x)$.

The Ekeland's variational principle guarantees the existence of such ε-solution where neither compactness nor convexity assumption on the underlying space is needed.

Theorem 9.1 (Strong Form of Ekeland's Variational Principle). *Let (X,d) be a complete metric space and $f: X \to \mathbb{R} \cup \{+\infty\}$ be a proper, bounded below and lower semicontinuous functional. Let $\varepsilon > 0$ and $\hat{x} \in X$ be given such that*

$$f(\hat{x}) \le \inf_X f + \varepsilon.$$

Then for a given $\lambda > 0$, there exists $\bar{x} \in X$ such that

(a) $f(\bar{x}) \le f(\hat{x})$
(b) $d(\hat{x},\bar{x}) \le \lambda$
(c) $f(\bar{x}) < f(x) + \frac{\varepsilon}{\lambda} d(x,\bar{x})$ *for all* $x \in X \setminus \{\bar{x}\}$.

Proof. For the sake of convenience, we set $d_\lambda(u,v) = (1/\lambda)d(u,v)$. Then d_λ is equivalent to d and (X,d_λ) is complete. Let us define a partial ordering $\preccurlyeq$ on X by

$$x \preccurlyeq y \quad \text{if and only if} \quad f(x) \le f(y) - \varepsilon d_\lambda(x,y).$$

It is easy to see that this ordering is (i) reflexive, that is, for all $x \in X$, $x \preccurlyeq x$; (ii) antisymmetric, that is, for all $x,y \in X$, $x \preccurlyeq y$ and $y \preccurlyeq x$ imply $x = y$; (iii) transitive, that is, for all $x,y,z \in X$, $x \preccurlyeq y$ and $y \preccurlyeq z$ imply $x \preccurlyeq z$.

We define a sequence $\{S_n\}$ of subsets of X as follows: Start with $x_1 = \hat{x}$ ($\hat{x}$ is the same as given in the statement of the theorem) and define

$$S_1 = \{x \in X : x \preccurlyeq x_1\}; \quad x_2 \in S_1 \text{ such that } f(x_2) \le \inf_{S_1} f + \frac{\varepsilon}{2},$$

$$S_2 = \{x \in X : x \preccurlyeq x_2\}; \quad x_3 \in S_2 \text{ such that } f(x_3) \le \inf_{S_2} f + \frac{\varepsilon}{2^2},$$

and inductively

$$S_n = \{x \in X : x \preccurlyeq x_n\}; \quad x_{n+1} \in S_n \text{ such that } f(x_{n+1}) \leq \inf_{S_n} f + \frac{\varepsilon}{2^n}.$$

Clearly, $S_1 \supset S_2 \supset S_3 \supset \ldots$. We claim that each S_n is closed.

Indeed, let $u_j \in S_n$ with $u_j \to u \in X$. Then $u_j \preccurlyeq x_n$ and so $f(u_j) \leq f(x_n) - \varepsilon d_\lambda(u_j, x_n)$. By taking limit and using the lower semicontinuity of f and the continuity of d and so the continuity of d_λ, we conclude that $u \in S_n$.

Now, we prove that the diameter of these sets S_n, $\delta(S_n) \to 0$ as $n \to \infty$.

Indeed, take an arbitrary point $u \in S_n$, On one hand, $u \preccurlyeq x_n$ implies that

$$f(u) \leq f(x_n) - \varepsilon d_\lambda(u, x_n). \tag{9.1}$$

On the other hand, we observe that u also belongs to S_{n-1}. So it is one of the points which entered in the competition when we picked x_n. Therefore,

$$f(x_n) \leq f(u) + \frac{\varepsilon}{2^{n-1}}. \tag{9.2}$$

From (9.1) and (9.2), we obtain

$$d_\lambda(u, x_n) \leq 2^{-n+1} \quad \text{for all } u \in S_n$$

which gives $\delta(S_n) \leq 2^{-n}$ and hence $\delta(S_n) \to 0$ as $n \to \infty$.

Since (X, d_λ) is complete and $\{S_n\}$ is a decreasing sequence of closed sets, by Cantor's Intersection Theorem 2.9, we infer that

$$\bigcap_{n=1}^{\infty} S_n = \{\bar{x}\}.$$

We still have to prove that this unique point $\bar{x}$ satisfies conditions (a)–(c).

Since $\bar{x} \in S_1$, we have

$$\bar{x} \preccurlyeq x_1 = \hat{x} \quad \text{if and only if} \quad f(\bar{x}) \leq f(\hat{x}) - \varepsilon d_\lambda(\bar{x}, \hat{x})$$

and so $f(\bar{x}) \leq f(\hat{x})$. Hence (a) is proved.

Now let $x \neq \bar{x}$. We can not have $x \preccurlyeq \bar{x}$, because otherwise x would belong to $\bigcap_{n=1}^{\infty} S_n$. So $x \not\preccurlyeq \bar{x}$, which means that

$$f(x) > f(\bar{x}) - \varepsilon d_\lambda(x, \bar{x})$$

and hence (c) is proved.

Finally, by writing

$$d_\lambda(\hat{x}, x_n) = d_\lambda(x_1, x_n) \leq \sum_{j=1}^{n-1} d_\lambda(x_j, x_{j+1}) \leq \sum_{j=1}^{n-1} 2^{-j}$$

and taking limit as $n \to \infty$, we obtain $d_\lambda(\hat{x}, \bar{x}) \leq 1$ and so $d(\hat{x}, \bar{x}) \leq \lambda$. This proves (b).

Remark 9.1. Let $S_r[x]$ be the closed sphere with center at x and radius $r > 0$. In this context, condition (b) of Theorem 9.1 can be written as $\bar{x} \in S_\lambda[\hat{x}]$.

Remark 9.2. Strong form of Ekeland's variational principle says that for $\lambda, \varepsilon > 0$ and $\hat{x}$ an ε-approximate solution of an optimization problem, there exists a new point $\bar{x}$ that is not worse than $\hat{x}$ and belongs to a λ-neighborhood of $\hat{x}$, and especially, $\bar{x}$ satisfies (c). Relation (c) says, in fact, that $\bar{x}$ minimizes globally $f(\cdot) + \left(\frac{\varepsilon}{\lambda}\right) d(\cdot, \bar{x})$, which is a Lipschitz perturbation of f.

Aubin and Frankowska [7] established the following form of Ekeland's variational principle which is equivalent to Theorem 9.1.

Theorem 9.2 (Strong Form of Ekeland's Variational Principle). *Let (X,d) be a complete metric space and $f : X \to \mathbb{R} \cup \{+\infty\}$ be a proper, bounded below and lower semicontinuous functional. Let $\hat{x} \in dom(f)$ and $e > 0$ be fixed. Then there exists $\bar{x} \in X$ such that*

(aa) $f(\bar{x}) - f(\hat{x}) + ed(\hat{x}, \bar{x}) \leq 0$
(bb) $f(\bar{x}) < f(x) + ed(x, \bar{x})$ *for all* $x \in X \setminus \{\bar{x}\}$.

Proof. For the sake of convenience, we set $d_e(u,v) = ed(u,v)$. Then d_e is equivalent to d and (X, d_e) is complete. For all $x \in X$, define

$$S(x) = \{y \in X : f(y) - f(x) + d_e(x,y) \leq 0\}.$$

We assume that $f(x) \neq +\infty$ for all $x \in X$. For all $x \in X$, we have $x \in S(x)$, and therefore, $S(x)$ is nonempty. So, we can let $y \in S(x)$. Then

$$f(y) - f(x) + d_e(x,y) \leq 0 \tag{9.3}$$

and also let $z \in S(y)$, then

$$f(z) - f(y) + d_e(y,z) \leq 0. \tag{9.4}$$

Adding (9.3) and (9.4), we obtain

$$f(z) - f(x) + d_e(x,z) \leq 0.$$

Therefore, $z \in S(x)$ which implies that $S(y) \subseteq S(x)$.

We claim that for every $x \in X$, $S(x)$ is closed. Indeed, let $\{x_n\}$ be a sequence in $S(x)$ such that $x_n \to x^* \in X$. Then

$$f(x_n) - f(x) + d_e(x, x_n) \leq 0.$$

By using the lower semicontinuity of f and continuity of d and hence continuity of d_e, we have

$$f(x^*) - f(x) + d_e(x,x^*) \leq \lim_{n\to\infty} f(x_n) - f(x) + \lim_{n\to\infty} d_e(x,x_n) \leq 0.$$

Therefore, $x^* \in S(x)$ and so $S(x)$ is closed for every $x \in X$.

For all $x \in \text{dom}(f)$, define $\mathcal{V}(x)$ by

$$\mathscr{V}(x) := \inf_{z \in S(x)} f(z).$$

For every $z \in S(x)$,

$$d_e(x,z) \leq f(x) - f(z) \leq f(x) - \mathscr{V}(x),$$

so that the diameter of $S(x)$, $\delta(S(x)) = \sup_{y,z \in S(x)} d_e(y,z)$, is not greater than $2\big(f(x) - \mathscr{V}(x)\big)$.

Define a sequence in the following manner: Fix $x_0 = \hat{x} \in X$; take $x_1 \in S(x_0)$ such that

$$f(x_1) \leq \mathscr{V}(x_0) + 2^{-1}.$$

Denote by x_2 any point in $S(x_1)$ such that

$$f(x_2) \leq \mathscr{V}(x_1) + 2^{-2}.$$

Continue in this manner, we obtain a sequence $\{x_n\}$ such that

$$x_{n+1} \in S(x_n) \quad \text{and} \quad f(x_{n+1}) \leq \mathscr{V}(x_n) + 2^{-(n+1)}.$$

Since $S(x_{n+1}) \subseteq S(x_n)$, we deduce that

$$\mathscr{V}(x_{n+1}) = \inf_{z \in S(x_{n+1})} f(z) \geq \inf_{z \in S(x_n)} f(z) = \mathscr{V}(x_n)$$

and thus

$$\mathscr{V}(x_n) \leq \mathscr{V}(x_{n+1}).$$

On the other hand, inequality $\mathscr{V}(y) \leq f(y)$ implies that

$$\mathscr{V}(x_{n+1}) \leq f(x_{n+1}) \leq \mathscr{V}(x_n) + 2^{-(n+1)} \leq \mathscr{V}(x_{n+1}) + 2^{-(n+1)}$$

and thus

$$0 \leq 2\left(f(x_{n+1}) - \mathscr{V}(x_{n+1})\right) \leq 2^{-n}.$$

It follows that

$$\delta(S(x_n)) \leq 2^{-n} \to 0 \quad \text{as } n \to \infty.$$

By Cantor's Intersection Theorem 2.9, there exists exactly one point $\bar{x} \in X$ such that

$$\bigcap_{n=0}^{\infty} S(x_n) = \{\bar{x}\}.$$

This implies that $\bar{x} \in S(x_0) = S(\hat{x})$, that is,

$$f(\bar{x}) - f(\hat{x}) + d_e(\hat{x},\bar{x}) \leq 0$$

and so (aa) holds.

Moreover, $\bar{x}$ also belongs to all $S(x_n)$ and, since $S(\bar{x}) \subseteq S(x_n)$ for all n, we have

$$S(\bar{x}) = \{\bar{x}\}.$$

It follows that $x \notin S(\bar{x})$ whenever $x \neq \bar{x}$ implying that

$$f(x) - f(\bar{x}) + d_e(\bar{x},x) > 0,$$

that is, (bb) holds.

Remark 9.3. Theorem 9.1 and Theorem 9.2 are equivalent.

Proof. Assume that Theorem 9.2 and the hypothesis of Theorem 9.1 hold. By (aa), we have $f(\bar{x}) \leq f(\hat{x}) - ed(\hat{x},\bar{x})$. Since $e > 0$, we obtain $f(\bar{x}) \leq f(\hat{x})$ and so (a) holds.

By hypothesis of Theorem 9.1, $f(\hat{x}) \leq \inf_X f + \varepsilon$, that is, $f(\hat{x}) - \inf_X f \leq \varepsilon$. Therefore, in particular, $f(\hat{x}) - f(\bar{x}) \leq \varepsilon$. Set $e = \frac{\varepsilon}{\lambda}$. Then by (aa) of Theorem 9.2, we have

$$d(\hat{x},\bar{x}) \leq \frac{\lambda}{\varepsilon}[f(\hat{x}) - f(\bar{x})] \leq \frac{\lambda}{\varepsilon}\varepsilon = \lambda$$

and (b) of Theorem 9.1 follows.

Conversely, suppose that Theorem 9.1 and the hypothesis of Theorem 9.2 hold. Set $\varepsilon = e\lambda$ and let $\hat{x} \in dom(f)$ and $e > 0$ be given. Take $f(\hat{x}) - \inf_X f \leq \varepsilon$, and consider

$$S_1 := \{x \in X : f(x) - f(\hat{x}) \leq -\varepsilon\}$$

and

$$S_2 := \{x \in X : f(x) - f(\hat{x}) + ed(\hat{x},x) \leq 0\}.$$

The lower semicontinuity of f implies that S_1 and S_2 are closed. Furthermore, $S_2 \neq \emptyset$ as $\hat{x} \in S_2$. By applying Theorem 9.1 for ε, λ and $S_1 \cup S_2$ instead of X, we obtain $\bar{x} \in S_1 \cup S_2$ such that (a), (b) and

$$(cc) \qquad f(x) - f(\bar{x}) + \frac{\varepsilon}{\lambda}d(x,\bar{x}) > 0 \quad \text{for all } x \in S_1 \cup S_2 \setminus \{\bar{x}\}$$

hold.

If $\bar{x} \in S_2$, then (aa) holds. If $\bar{x} \in S_1$, then we have

$$f(\bar{x}) \leq f(\hat{x}) - \varepsilon.$$

Adding this inequality and (b) multiplied by e, we obtain

$$f(\bar{x}) - f(\hat{x}) + ed(\hat{x},\bar{x}) \leq 0,$$

that is, $\bar{x} \in S_2$. So (aa) holds.

To show (bb), it is sufficient to check (cc) is satisfied also for $x \notin S_1 \cup S_2$. By the definition of S_2, $x \notin S_2$ means $f(x) - f(\hat{x}) + ed(\hat{x},x) > 0$. Adding this inequality and (aa), we get

$$f(x) - f(\bar{x}) + ed(x,\bar{x}) \geq 0,$$

that is, (bb) holds.

We now present, so called, the weak formulation of Ekeland's variational principle.

Corollary 9.1 (Weak Form of Ekeland's Variational Principle). *Let (X,d) be a complete metric space and $f : X \to \mathbb{R} \cup \{+\infty\}$ be a proper, bounded below and lower*

semicontinuous functional. Then for any given $\varepsilon > 0$, there exists $\bar{x} \in X$ such that

$$f(\bar{x}) \leq \inf_X f + \varepsilon$$

and

$$f(\bar{x}) < f(x) + \varepsilon d(x,\bar{x}) \quad \textit{for all } x \in X \textit{ with } x \neq \bar{x}. \tag{9.5}$$

Proof. It follows from the fact that there always exists some point $\hat{x} \in X$ such that

$$f(\hat{x}) \leq \inf_X f + \varepsilon.$$

Then by Theorem 9.1, there exists $\bar{x} \in X$ (by taking $\lambda = 1$) such that

$$f(\bar{x}) \leq f(\hat{x}) \leq \inf_X f + \varepsilon$$

and

$$f(\bar{x}) < f(x) + \varepsilon d(x,\bar{x}) \quad \text{for all } x \in X \text{ with } x \neq \bar{x}.$$

The property of Ekeland's variational principle for proper but extended real-valued lower semicontinuous and bounded below functions on a metric space characterizes compactness of the metric space.

Theorem 9.3 (Converse of Ekeland's Variational Principle). *A metric space (X,d) is complete if for every functional $f : X \to \mathbb{R} \cup \{+\infty\}$ which is proper, bounded below and lower semicontinuous on X and for every given $\varepsilon > 0$, there exists $\bar{x} \in X$ such that*

$$f(\bar{x}) \leq \inf_X f + \varepsilon$$

and

$$f(\bar{x}) \leq f(x) + \varepsilon d(x,\bar{x}) \quad \textit{for all } x \in X.$$

Proof. Consider a Cauchy sequence $\{x_n\}$ in X and define a functional $f : X \to [0,+\infty)$ by

$$f(x) = \lim_{n\to\infty} d(x_n,x) \quad \text{for all } x \in X.$$

Then f is well-defined, nonnegative and continuous. Since $\{x_n\}$ is Cauchy, we have

$$f(x_m) = \lim_{n\to\infty} d(x_n,x_m) \to 0 \quad \text{as } m \to \infty.$$

Therefore, $\inf_X f = 0$.

For given $0 < \varepsilon < 1$, choose $\bar{x} \in X$ such that $f(\bar{x}) \leq \varepsilon$ and

$$f(\bar{x}) \leq f(x) + \varepsilon d(x,\bar{x}) \quad \text{for all } x \in X. \tag{9.6}$$

Letting $x = x_n$ in (9.6) and taking limit as $n \to \infty$, we obtain $f(\bar{x}) \leq \varepsilon f(\bar{x})$ and so that $f(\bar{x}) = 0$. This implies that $\{x_n\}$ converges to $\bar{x}$.

9.2 Applications to Fixed Point Theorems

As a first application of Ekeland's variational principle, we prove the well-known Banach contraction theorem.

Theorem 9.4 (Banach Contraction Theorem). *Let X be a complete metric space and $T : X \to X$ be a contraction mapping. Then T has a unique fixed point in X.*

Proof. Consider the functional $f : X \to [0,\infty)$ defined by

$$f(x) = d(x,T(x)) \quad \text{for all } x \in X.$$

Then f is bounded below and continuous on X. Choose ε such that $0 < \varepsilon < 1 - \alpha$, where α is the Lipschitz constant. By Corollary 9.1, there exists $\bar{x} \in X$ (depending on ε) such that

$$f(\bar{x}) \le f(x) + \varepsilon d(x,\bar{x}) \quad \text{for all } x \in X.$$

Putting $x = T(\bar{x})$, we have

$$\begin{aligned} d(\bar{x},T(\bar{x})) &\le d(T(\bar{x}),T(T(\bar{x}))) + \varepsilon d(\bar{x},T(\bar{x})) \\ &\le \alpha d(\bar{x},T(\bar{x})) + \varepsilon d(\bar{x},T(\bar{x})) \\ &= (\alpha+\varepsilon) d(\bar{x},T(\bar{x})). \end{aligned}$$

If $\bar{x} \neq T(\bar{x})$, then we obtain $1 \le \alpha + \varepsilon$, which contradicts to our assumption that $\alpha + \varepsilon < 1$. Therefore, we have $\bar{x} = T(\bar{x})$. The uniqueness of $\bar{x}$ can be proved as in the original proof of the Banach contraction theorem.

As a second application of Ekeland's variational principle, we prove Caristi's fixed point theorem.

Theorem 9.5 (Caristi's Fixed Point Theorem). *Let X be a complete metric space and $T : X \to X$ be a mapping such that*

$$d(x,T(x)) + \varphi(T(x)) \le \varphi(x) \quad \textit{for all } x \in X, \tag{9.7}$$

where $\varphi : X \to \mathbb{R} \cup \{+\infty\}$ is a proper, lower semicontinuous and bounded below functional. Then there exists $\bar{x} \in X$ such that $T(\bar{x}) = \bar{x}$ and $\varphi(\bar{x}) < \infty$.

Proof. By using Corollary 9.1 with $\varepsilon = 1$, we obtain $\bar{x} \in X$ such that

$$\varphi(\bar{x}) < \varphi(x) + d(x,\bar{x}) \quad \text{for all } x \in X \setminus \{\bar{x}\}. \tag{9.8}$$

We claim that $\bar{x} = T(\bar{x})$. Otherwise all $y = T(\bar{x}) \in X$ are such that $y \neq \bar{x}$. Then from (9.7) and (9.8), we have

$$d(y,\bar{x}) + \varphi(y) \le \varphi(\bar{x}) \quad \text{and} \quad \varphi(\bar{x}) < \varphi(y) + d(y,\bar{x})$$

which cannot hold simultaneously.

The following result, established by Park [70], provides some equivalent formulations of Ekeland's variational principle.

Theorem 9.6 (Equivalent Formulations of Ekeland's Variational Principle). *Let X be a complete metric space and $f : X \to \mathbb{R} \cup \{+\infty\}$ be a proper, bounded below and lower semicontinuous functional. Let $\varepsilon > 0$ and $\hat{x} \in X$ be given such that*

$$f(\hat{x}) \leq \inf_X f + \varepsilon.$$

Then the following statements are equivalent:

(a) For a given $\lambda > 0$, there exists $\bar{x} \in \overline{S} := \overline{S_\lambda(\hat{x})}$ such that $f(\bar{x}) \leq f(\hat{x})$ and

$$f(\bar{x}) < f(x) + \frac{\varepsilon}{\lambda} d(x, \bar{x}) \quad \text{for all } x \neq \bar{x}.$$

(b) If $T : \overline{S} \rightrightarrows X$ is a set-valued map satisfying the following condition:

$$\forall\, x \in \overline{S} \setminus T(x), \ \exists\, y \in X \setminus \{x\} \text{ such that } f(y) \leq f(x) - \frac{\varepsilon}{\lambda} d(y, x).$$

Then T has a fixed point $\bar{x} \in \overline{S}$, that is, $\bar{x} \in T(\bar{x})$ such that $f(\bar{x}) \leq f(\hat{x})$.

(c) If $\varphi : \overline{S} \to X$ is a single-valued map such that

$$f(\varphi(x)) \leq f(x) - \frac{\varepsilon}{\lambda} d(\varphi(x), x) \quad \text{for all } x \in \overline{S},$$

then φ has a fixed point $\bar{x} \in \overline{S}$ such $f(\bar{x}) \leq f(\hat{x})$.

(d) If $T : \overline{S} \rightrightarrows X$ is a set-valued map such that for all $x \in \overline{S}$, $T(x) \neq \emptyset$, and for any $y \in T(x)$, we have

$$f(y) \leq f(x) - \frac{\varepsilon}{\lambda} d(y, x),$$

then T has a stationary point $\bar{x} \in \overline{S}$, that is, $T(\bar{x}) = \{\bar{x}\}$ such that $f(\bar{x}) \leq f(\hat{x})$.

Proof. $(a) \Rightarrow (b)$: By (a), there exists a point $\bar{x} \in \overline{S}$ such that

$$f(\bar{x}) < f(x) + \varepsilon \lambda^{-1} d(x, \bar{x}) \quad \text{for all } x \neq \bar{x}. \tag{9.9}$$

Suppose that $\bar{x} \notin T(\bar{x})$, then by assumption, there exists a $y \neq \bar{x}$ such that

$$f(y) \leq f(\bar{x}) - \varepsilon \lambda^{-1} d(y, \bar{x}).$$

By putting $y = x$, we reach to a contradiction of (9.9).

$(b) \Rightarrow (c)$: Let $T : \overline{S} \rightrightarrows X$ be defined as $T(x) = \{\varphi(x)\}$ for all $x \in \overline{S}$. For any $x \in \overline{S}$, let $x \neq \varphi(x)$; Otherwise, every point of $\overline{S}$ is a fixed point of φ. Then $x \in \overline{S} \setminus T(x)$ and also $y = \varphi(x) \in X \setminus \{x\}$ satisfying

$$f(y) \leq f(x) - \varepsilon \lambda^{-1} d(y, x).$$

Then from (b), there exists $\bar{x} \in T(\bar{x}) = \{\varphi(\bar{x})\}$ such that $f(\bar{x}) \leq f(\hat{x})$.

$(c) \Rightarrow (d)$: By (c), any single-valued map φ on $\{T(x) : x \in \overline{S}\}$ has a fixed point. Assume to the contrary that T has no stationary point. Consider a map φ such that for any $x \in \overline{S}$, $x \in T(x)$ implies $\varphi(x) \in T(x) \setminus \{x\}$. Then such φ can not have a fixed point, which is a contradiction.

$(d) \Rightarrow (a)$: Define a set-valued map $T : \overline{S} \rightrightarrows X$ by

$$T(x) = \{y \in X : f(y) \leq f(x) - \varepsilon\lambda^{-1}d(y,x)\}.$$

Then for all $x \in \overline{S}$, $T(x) \neq \emptyset$ and by (d) T has a stationary point $\bar{x}$ which satisfies condition (a).

Definition 9.1. Let (X,d) be a metric space. For any $x,y \in X$, the *segment* between x and y is defined by

$$[x,y] = \{z \in X : d(x,z) + d(z,y) = d(x,y)\}. \tag{9.10}$$

Definition 9.2. Let (X,d) be a metric space. A mapping $T : X \to X$ is said to be a *directional contraction* if

(i) T is continuous, and
(ii) there exists $\alpha \in (0,1)$ such that for any $x \in X$ with $T(x) \neq x$, there exists $z \in [x,T(x)] \setminus \{x\}$ such that

$$d(T(x),T(z)) \leq \alpha d(x,z).$$

Theorem 9.7 (Clarke's Fixed Point Theorem). *Let (X,d) be a complete metric space and $T : X \to X$ be a directional contraction mapping. Then T has a fixed point.*

Proof. Define the functional $f : X \to [0,+\infty)$ by

$$f(x) = d(x,T(x)) \quad \text{for all } x \in X.$$

Then T is continuous and bounded below. Applying Corollary 9.1 to f with $\varepsilon \in (0,1-\alpha)$, we conclude that there exists $\bar{x} \in X$ such that

$$f(\bar{x}) \leq f(x) + \varepsilon d(x,\bar{x}) \quad \text{for all } x \in X. \tag{9.11}$$

If $T(\bar{x}) = \bar{x}$, then we are done. Otherwise, by the definition of a directional contraction mapping, there exists a point $z \neq \bar{x}$ with $z \in [\bar{x},T(\bar{x})]$, that is,

$$d(\bar{x},z) + d(z,T(\bar{x})) = d(\bar{x},T(\bar{x})) = f(\bar{x}) \tag{9.12}$$

satisfying

$$d(T(\bar{x}),T(\bar{z})) \leq \alpha d(\bar{x},z). \tag{9.13}$$

Setting $x = z$ in (9.11), we get

$$f(\bar{x}) \leq f(z) + \varepsilon d(z,\bar{x}).$$

By using the above inequality and (9.12), we obtain

$$d(\bar{x},z) + d(z,T(\bar{x})) \leq d(z,T(z)) + \varepsilon d(z,\bar{x})$$

or

$$d(\bar{x},z) \leq d(z,T(z)) - d(z,T(\bar{x})) + \varepsilon d(z,\bar{x}). \tag{9.14}$$

By the triangle inequality and (9.13), we obtain

$$d(z,T(z)) - d(z,T(\bar{x})) \leq d(T(\bar{x}),T(z)) \leq \alpha d(\bar{x},z). \tag{9.15}$$

Combining (9.14) and (9.15), we have

$$d(\bar{x},z) \leq (\alpha+\varepsilon)d(\bar{x},z),$$

a contradiction.

Remark 9.4. It is clear that every contraction mapping is a directional contraction. Therefore, Clarke's fixed point theorem is more general than Banach contraction theorem.

In the following example, we show that Theorem 9.7 is applicable, however, Theorem 9.4 is not.

Example 9.1. Let $X = \mathbb{R}^2$ be a metric space with the metric

$$d(x,y) = |x_1 - y_1| + |x_2 - y_2| \quad \text{for all } x = (x_1,x_2),\ y = (y_1,y_2) \in \mathbb{R}^2.$$

A segment between two points (a_1,a_2) and (b_1,b_2) consists of the closed rectangle having the two points as diagonally opposite corners. Define

$$T(x) = \left(\frac{3x_1}{2} - \frac{x_2}{3},\ x_1 + \frac{x_2}{3}\right) \quad \text{for all } x = (x_1,x_2) \in \mathbb{R}^2.$$

Then T is a directional contraction mapping. Indeed, if $y = (y_1,y_2) = T(x) \neq x = (x_1,x_2)$. Then $y_2 \neq x_2$ (for otherwise we will also have $y_1 = x_1$). Now the set $[x,y]$ contains points of the form (x_1,t) with t arbitrarily close to x_2 but not equal to x_2. For such points, we have

$$d\left(T(x_1,t),T(x_1,x_2)\right) = \frac{2}{3}d\left((x_1,t),(x_1,x_2)\right).$$

Hence T is a directional contraction mapping. We can easily check that the fixed points of T are all points of the form $(x,3x/2)$. Since T has more than one fixed point clearly then Banach contraction theorem does not apply to this mapping.

9.3 Applications to Optimization

We derive the following existence result for a solution of an optimization problem without compactness and convexity assumptions on the underlying space.

Theorem 9.8 (Takahashi's Minimization Theorem). *Let X be a complete metric space and $f : X \to \mathbb{R} \cup \{+\infty\}$ be a proper, bounded below and lower semicontinuous functional. Suppose that, for each $\hat{x} \in X$ with $\inf_X f < f(\hat{x})$, there exists $z \in X$ such that $z \neq \hat{x}$ and*

$$f(z) + d(\hat{x},z) \leq f(\hat{x}).$$

Then there exists $\bar{x} \in X$ such that $f(\bar{x}) = \inf_{x \in X} f(x)$, that is, $\bar{x}$ is a solution of OP.

Proof. Assume to the contrary that $\inf_{x\in X} f(x) < f(y)$ for all $y \in X$ and let $\hat{x} \in X$ with $f(\hat{x}) < +\infty$. We define inductively a sequence $\{x_n\}$ in X, starting with $x_1 = \hat{x}$. Suppose that $x_n \in X$ is known. Then choose $x_{n+1} \in S_{n+1}$ such that

$$S_{n+1} = \{x \in X : f(x) \le f(x_n) - d(x_n, x)\}$$

and

$$f(x_{n+1}) \le \frac{1}{2}\left\{\inf_{x\in S_{n+1}} f(x) + f(x_n)\right\}. \tag{9.16}$$

We claim that $\{x_n\}$ is a Cauchy sequence. Indeed, if $m > n$, then

$$\begin{aligned} d(x_n, x_m) &\le \sum_{j=n}^{m-1} d(x_j, x_{j+1}) \\ &\le \sum_{j=n}^{m-1} \{f(x_j) - f(x_{j+1})\} \\ &= f(x_n) - f(x_m). \end{aligned} \tag{9.17}$$

Since $\{f(x_n)\}$ is a decreasing sequence and the functional f is bounded below, there exists $\varepsilon > 0$ such that

$$f(x_n) - f(x_m) < \varepsilon \quad \text{for all } m > n.$$

Therefore from (9.17), we have

$$d(x_n, x_m) < \varepsilon \quad \text{for all } m > n$$

and hence $\{x_n\}$ is a Cauchy sequence in X. Since X is a complete metric space, there exists $\bar{x} \in X$ such that $x_n \to \bar{x}$. Then, if $m \to \infty$ in (9.17), the lower semicontinuity of f and continuity of d imply that

$$d(x_n, \bar{x}) \le f(x_n) - \lim_{m\to\infty} f(x_m) \le f(x_n) - f(\bar{x}). \tag{9.18}$$

On the other hand, by hypothesis, there exists a $z \in X$ such that $z \neq \bar{x}$ and

$$f(z) + d(\bar{x}, z) \le f(\bar{x}). \tag{9.19}$$

By using (9.18) and (9.19), we have

$$\begin{aligned} f(z) &\le f(\bar{x}) - d(\bar{x}, z) \\ &\le f(\bar{x}) - d(\bar{x}, z) + f(x_n) - f(\bar{x}) - d(x_n, \bar{x}) \\ &= f(x_n) - \{d(x_n, \bar{x}) + d(\bar{x}, z)\} \\ &\le f(x_n) - d(x_n, z). \end{aligned}$$

Consequently, $z \in S_{n+1}$ for all $n \in \mathbb{N}$. Using (9.16), we obtain

$$2f(x_{n+1}) - f(x_n) \le \inf_{x\in S_{n+1}} f(x) \le f(z).$$

Hence

$$f(\bar{x}) \leq \lim_{n\to\infty} f(x_n) \leq f(z) \leq f(\bar{x}) - d(\bar{x},z) < f(\bar{x})$$

which is a contradiction. Therefore, there exists $\bar{x} \in X$ such that $f(\bar{x}) = \inf_{x\in X} f(x)$.

Remark 9.5. Takahashi's Minimization Theorem 9.8 and Ekeland's Variational Principle Theorem 9.1 are equivalent.

Proof. We first prove Theorem 9.1 by using Theorem 9.8. Let

$$X_0 = \{x \in X : f(x) \leq f(\hat{x}) - \varepsilon d_\lambda(\hat{x},x)\}.$$

Since $\hat{x} \in X_0$, X_0 is nonempty. By lower semicontinuity of f and continuity of d_λ, X_0 is closed. Further, for each $x \in X_0$,

$$\varepsilon d_\lambda(\hat{x},x) \leq f(\hat{x}) - f(x) \leq f(\hat{x}) - \inf_{y\in X} f(y) \leq \varepsilon$$

and hence $d_\lambda(\hat{x},x) \leq 1$ and thus $d(\hat{x},x) \leq \lambda$. We also have $f(x) \leq f(\hat{x})$.

Assume to the contrary of conclusion (c) in Theorem 9.1 that for every $x \in X_0$, there exists $y \in X$ such that $y \neq x$ and $f(y) \leq f(x) - \varepsilon d_\lambda(y,x)$. Then

$$\begin{aligned} \varepsilon d_\lambda(\hat{x},y) &\leq \varepsilon d_\lambda(\hat{x},x) + \varepsilon d_\lambda(x,y) \\ &\leq f(\hat{x}) - f(x) + f(x) - f(y) \\ &= f(\hat{x}) - f(y) \end{aligned}$$

and hence $y \in X_0$. Then by Theorem 9.8, there exists $\bar{x} \in X$ such that $f(\bar{x}) = \inf_{x\in X_0} f(x)$. This is a contradiction of the hypothesis that there exists $y_0 \in X_0$ with $f(y_0) < f(\bar{x})$.

Now we prove Theorem 9.8 by using Theorem 9.1. By Theorem 9.1, for any given $\varepsilon > 0$, there exists $\bar{x} \in X$ such that

$$f(\bar{x}) < f(x) + \varepsilon d(x,\bar{x}) \quad \text{for all } x \in X \text{ with } x \neq \bar{x}. \tag{9.20}$$

We claim that $f(\bar{x}) = \inf_{x\in X} f(x)$.

Assume to the contrary that there exists $w \in X$ such that $f(w) > \inf_{x\in X} f(x)$. By hypothesis, there exists $z \in X$ such that $z \neq w$ and

$$f(z) + d(w,z) \leq f(w)$$

contradicting (9.20). Hence $f(\bar{x}) = \inf_{x\in X} f(x)$.

Remark 9.6. Takahashi [86] also derived Caristi's fixed point theorem [22], Nadler's fixed point theorem for set-valued maps [64] and Fan's minimization theorem [37] by using Theorem 9.8.

9.4 Equilibrium Problems and Extended Ekeland's Variational Principle

9.4.1 Equilibrium Problems

Investigations of equilibrium states of a system play a central role in such diverse fields as economics, mechanics, biology and social sciences. Now there are a number of general mathematical problems which were suggested for modeling and studying various kinds of equilibria. Many researchers were / are considering these problems in order to obtain existence and uniqueness results and to propose solution methods. The mathematical equilibrium problem (in short, EP), which is to find an element $\bar{x}$ of a set K such that

$$F(\bar{x},y) \geq 0 \quad \text{for all } y \in K, \tag{9.21}$$

where $F : K \times K \to \mathbb{R}$ is a bifunction such that $F(x,x) = 0$ for all $x \in K$, seems the most general problem and includes other equilibrium type ones such as optimization, saddle point, fixed point, complementarity and variational inequality ones. In this general form, EP was first considered by Nikaido and Isoda [68] as an auxiliary problem to establish existence results for Nash's equilibrium points in non-cooperative games. This transformation allows one to extend various iterative methods, which were proposed for saddle point problems, for the case of EP. In the theory of EP's, the key contribution was made by Ky Fan [38], whose new existence results contained the original techniques which became a basis for most further existence theorems in topological spaces. Within the context of calculus of variations, motivated mainly by the works of Stampacchia [81], there arises the work of Brézis, Niremberg and Stampacchia [19] establishing a more general result than that in [38]. The equilibrium problem is a unified model of several problems, for example, optimization problem, saddle point problem, Nash equilibrium problem, variational inequality problem, nonlinear complementarity problem, fixed point problem, etc. In the last decade, it emerges as a new direction of research in nonlinear analysis, optimization, optimal control, game theory, mathematical economics, etc. Most of the results on the existence of solutions of equilibrium problems are studied in the setting of topological vector spaces by using some kind of fixed point (Fan-Browder type fixed point) theorem or KKM type theorem. After the work of Blum and Otteli [14], many mathematicians have started to study the EP again. For further details on equilibrium problems, we refer to [3, 10, 11, 12, 14, 19, 24, 25, 40, 41, 47, 89] and the references therein.

Example 9.2. (a) **Minimization Problem.** Let K be a nonempty set and $\varphi : K \to \mathbb{R}$ be a real-valued function. The *minimization problem* (in short, MP) is to find $\bar{x} \in K$ such that

$$\varphi(\bar{x}) \leq \varphi(y) \quad \text{for all } y \in K. \tag{9.22}$$

If we set $F(x,y) = \varphi(y) - \varphi(x)$ for all $x,y \in K$, then MP is equivalent to EP.

(b) **Saddle Point Problem.** Let K_1 and K_2 be nonempty sets and $L : K_1 \times K_2 \to \mathbb{R}$ be a real-valued bifunction. The *saddle point problem* (in short, SPP) is to find $(\bar{x}_1,\bar{x}_2) \in K_1 \times K_2$ such that

$$L(\bar{x}_1,y_2) \leq L(\bar{x}_1,\bar{x}_2) \leq L(y_1,\bar{x}_2) \quad \text{for all } (y_1,y_2) \in K_1 \times K_2. \tag{9.23}$$

Set $K := K_1 \times K_2$ and define $F : K \times K \to \mathbb{R}$ by

$$F((x_1,x_2),(y_1,y_2)) = L(y_1,x_2) - L(x_1,y_2) \tag{9.24}$$

for all $(x_1,x_2),(y_1,y_2) \in K_1 \times K_2$. Then SPP coincides with EP.

(c) **Nash Equilibrium Problem.** Let I be a finite set of players. For each $i \in I$, let K_i be the strategy set of the ith player. Let $K := \prod_{i\in I} K_i$. For each player $i \in I$, let $\varphi_i : K \to \mathbb{R}$ be the loss function of the ith player, depending on the strategies of all players. For $x = (x_1,x_2,\ldots,x_n) \in K$, we define $x^i = (x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)$. The *Nash equilibrium problem* (in short, NEP) [65, 66] is to find $\bar{x} \in K$ such that for each $i \in I$,

$$\varphi_i(\bar{x}) \le \varphi_i(\bar{x}^i, y_i) \quad \text{for all } y_i \in K_i. \tag{9.25}$$

Define

$$F(x,y) = \sum_{i=1}^{n} \left(\varphi_i(x^i,y_i) - \varphi_i(x)\right).$$

Then NEP is same as EP.

(d) **Fixed Point Problem.** Let K be a nonempty subset of $\mathbb{R}^n$ and $\varphi : K \to K$ be a given mapping. The *fixed point problem* (in short, FPP) is to find $\bar{x} \in K$ such that $\varphi(\bar{x}) = \bar{x}$.

Setting $F(x,y) = \langle x - \varphi(x), y - x\rangle$, where $\langle .,. \rangle$ denotes the inner product on $\mathbb{R}^n$. Then $\bar{x}$ is a solution of FPP if and only if it is a solution of EP.

(e) **Variational Inequality Problem.** Let K be a nonempty subset of $\mathbb{R}^n$. Let $\Phi : K \to \mathbb{R}^n$ be a mapping. The *variational inequality problem* (in short, VIP) is to find $\bar{x} \in K$ such that

$$\langle \Phi(\bar{x}), y - \bar{x}\rangle \ge 0 \quad \text{for all } y \in K, \tag{9.26}$$

where $\langle .,. \rangle$ denotes the inner product on $\mathbb{R}^n$. We set $F(x,y) = \langle \Phi(x), y - x\rangle$ for all $x,y \in K$. Then VIP is equivalent to EP. For further details on variational inequality problems in the setting of finite dimensional spaces, we refer to [36].

9.4.2 Extended Ekeland's Variational Principle

We present the following equilibrium version of Ekeland's variational principle. It is called extended Ekeland's variational principle.

Theorem 9.9 (Extended Ekeland's Variational Principle). *Let (X,d) be a complete metric space. Let $F : X \times X \to \mathbb{R} \cup \{+\infty\}$ be a function such that it is lower semicontinuous in the second argument. Assume that the following conditions hold:*

(i) $F(x,x) = 0$ for all $x \in X$;
(ii) $F(x,y) \le F(x,z) + F(z,y)$ for all $x,y,z \in X$.

Then for every $e > 0$ and for every $\hat{x} \in X$, there exists $\bar{x} \in X$ such that

(aaa) $F(\hat{x},\bar{x}) + ed(\hat{x},\bar{x}) \le 0$
(bbb) $F(\bar{x},x) + ed(\bar{x},x) > 0$ *for all* $x \ne \bar{x}$.

Proof. For the sake of convenience, we set $d_e(u,v) = ed(u,v)$. Then d_e is equivalent to d and (X,d_e) is complete. For all $x \in X$, define

$$S(x) = \{y \in X : F(x,y) + d_e(x,y) \leq 0\}.$$

The lower semicontinuity of $F(x,\cdot)$ and continuity of d and hence continuity of d_e imply that $S(x)$ is closed for every $x \in X$. By condition (i), $x \in S(x)$ for all $x \in X$, and therefore, $S(x)$ is nonempty. So, we can assume $y \in S(x)$, that is, $F(x,y) + d_e(x,y) \leq 0$ and also let $z \in S(y)$, then $F(y,z) + d_e(z,y) \leq 0$. Adding these inequalities and using condition (ii), we obtain

$$0 \geq F(x,y) + d_e(x,y) + F(y,z) + d_e(z,y) \geq F(x,z) + d_e(x,z).$$

Therefore, $z \in S(x)$ which implies that $S(y) \subseteq S(x)$.

Define

$$\mathscr{V}(x) := \inf_{z \in S(x)} F(x,z).$$

For every $z \in S(x)$,

$$d_e(x,z) \leq -F(x,z) \leq \sup_{z \in S(x)} \big(-F(x,z)\big) = -\inf_{z \in S(x)} F(x,z) = -\mathscr{V}(x),$$

that is, $d_e(x,z) \leq -\mathscr{V}(x)$ for all $z \in S(x)$. In particular, if $x_1, x_2 \in S(x)$,

$$d_e(x_1,x_2) \leq d_e(x,x_1) + d_e(x,x_2) \leq -\mathscr{V}(x) - \mathscr{V}(x) = -2\mathscr{V}(x)$$

implying that the diameter of $S(x)$, $\delta(S(x)) = \sup_{y,z \in S(x)} d_e(y,z) \leq -2\mathscr{V}(x)$ for all $x \in X$.

Define a sequence in the following manner: Fix $x_0 = \hat{x} \in X$; $x_1 \in S(x_0)$ exists such that

$$F(x_0,x_1) \leq \mathscr{V}(x_0) + 2^{-1}.$$

Denote by x_2 any point in $S(x_1)$ such that

$$F(x_1,x_2) \leq \mathscr{V}(x_1) + 2^{-2}.$$

Continue in this manner, we obtain a sequence $\{x_n\}$ of points of X such that

$$x_{n+1} \in S(x_n) \quad \text{and} \quad F(x_n,x_{n+1}) \leq \mathscr{V}(x_n) + 2^{-(n+1)}.$$

Since $S(x_{n+1}) \subseteq S(x_n)$, we deduce that

$$\begin{aligned}
\mathscr{V}(x_{n+1}) &= \inf_{y \in S(x_{n+1})} F(x_{n+1},y) \geq \inf_{y \in S(x_n)} F(x_{n+1},y) \\
&\geq \inf_{y \in S(x_n)} \Big(F(x_n,y) - F(x_n,x_{n+1})\Big) \\
&\geq \inf_{y \in S(x_n)} \Big(F(x_n,y)\Big) - F(x_n,x_{n+1}) \\
&= \mathscr{V}(x_n) - F(x_n,x_{n+1}).
\end{aligned}$$

Therefore, $\mathscr{V}(x_{n+1}) \geq \mathscr{V}(x_n) - F(x_n,x_{n+1})$ and

$$-\mathscr{V}(x_n) \leq -F(x_n,x_{n+1}) + 2^{-(n+1)} \leq (\mathscr{V}(x_{n+1}) - \mathscr{V}(x_n)) + 2^{-(n+1)},$$

that entails $0 \leq \mathscr{V}(x_{n+1}) + 2^{-(n+1)}$. It follows that

$$\delta(S(x_n)) \leq -2\mathscr{V}(x_n) \leq 2 \cdot 2^{-n} \to 0 \quad \text{as } n \to \infty.$$

Since $\{S(x_n)\}$ is the family of closed sets and $S(x_{n+1}) \subseteq S(x_n)$, by Cantor's Intersection Theorem 2.9, there exists exactly one point $\bar{x} \in X$ such that

$$\bigcap_{n=0}^{\infty} S(x_n) = \{\bar{x}\}.$$

This implies that $\bar{x} \in S(x_0) = S(\hat{x})$, that is,

$$F(\hat{x},\bar{x}) + d_e(\hat{x},\bar{x}) \leq 0$$

and so (aaa) holds.

Moreover, $\bar{x}$ also belongs to all $S(x_n)$ and, since $S(\bar{x}) \subseteq S(x_n)$ for all n, we have $S(\bar{x}) = \{\bar{x}\}$. It follows that $x \notin S(\bar{x})$ whenever $x \neq \bar{x}$ implying that

$$F(\bar{x},x) + d_e(\bar{x},x) > 0,$$

that is, (bbb) holds.

Remark 9.7.

(i) Any function $F(x,y) = g(y) - g(x)$ trivially satisfies condition (ii), but there are other functions, not of this form, that fall into the framework of Theorem 9.9. Take, for instance, the function

$$F(x,y) = \begin{cases} e^{-\|x-y\|} + 1 + g(y) - g(x) & \text{if } x \neq y \\ 0 & \text{if } x = y, \end{cases}$$

where g is a bounded below and lower semicontinuous function.

(ii) Assumptions (i) and (ii) of Theorem 9.9 are satisfied for instance if $F(x,y) = \varphi(\Phi(y) - \Phi(x))$, where Φ maps X into a topological vector space V and $\varphi : X \to \mathbb{R} \cup \{+\infty\}$ is subadditive on X and satisfying $\varphi(0) = 0$.

Remark 9.8. The above theorem also holds if we replace the underlying set X by a closed subset K of a complete metric space X. In this case, with $X = \mathbb{R}^n$ Theorem 9.9 is established by Bianchi et al. [12].

Remark 9.9. Condition (ii) of Theorem 9.9 implies the cyclic monotonicity of $-F$ that extends a similar definition given for mappings in the framework of variational inequalities (see [73]): for every $x_1,x_2,\ldots,x_n \in X$ we have

$$\sum_{i=1}^{n} F(x_i,x_{i+1}) \geq 0, \tag{9.27}$$

where $x_{n+1} = x_1$. Indeed, taking $x = z$ in (ii), by (i) we get $F(x,y) + F(y,x) \geq 0$, and above inequality holds for $n = 2$. By induction, assuming that (9.27) holds for n, from (ii) the following inequalities hold:

$$\sum_{i=1}^{n+1} F(x_i,x_{i+1}) = \sum_{i=1}^{n-1} F(x_i,x_{i+1}) + F(x_n,x_{n+1}) + F(x_{n+1},x_1)$$
$$\geq \sum_{i=1}^{n-1} F(x_i,x_{i+1}) + F(x_n,x_1) \geq 0.$$

In particular, the cyclic monotonicity of $-F$ implies the monotonicity of $-F$.

Definition 9.3. Let K be a nonempty set, $F : K \times K \to \mathbb{R}$ be a bifunction and $\varepsilon > 0$ be given. The element $\bar{x} \in K$ is said to be an *ε-solution* of EP if

$$F(\bar{x},y) \geq -\varepsilon d(\bar{x},y) \quad \text{for all } y \in K. \tag{9.28}$$

It is called *strictly ε-solution* of EP if the inequality (9.28) is strict for all $x \neq y$.

Remark 9.10.

(i) Notice that (bbb) gives the existence of a strict ε-solution of EP for every $\varepsilon > 0$. Moreover, by conditions (i) and (ii) of Theorem 9.9, we have $F(\bar{x},\hat{x}) + F(\hat{x},\bar{x}) \geq F(\bar{x},\bar{x}) = 0$ and so $F(\bar{x},\hat{x}) \geq -F(\hat{x},\bar{x})$. It follows from (aaa) that

$$F(\bar{x},\hat{x}) \geq \varepsilon d(\bar{x},\hat{x}),$$

"localizing", in certain sense, the position of the $\bar{x}$.

(ii) Let $-\varepsilon := \inf_{x \in X} F(\hat{x},x)$. By replacing e by $1/n$, we obtain from Theorem 9.9 for arbitrary $n \in \mathbb{N}$, the existence of $\bar{x} \in X$ such that

$$F(\hat{x},\bar{x}) + \frac{1}{n} d(\hat{x},\bar{x}) \leq 0$$

and

$$F(\bar{x},x) + \frac{1}{n} d(\bar{x},x) \geq 0. \tag{9.29}$$

The first inequality implies in particular that

$$F(\hat{x},\bar{x}) \leq 0 \tag{9.30}$$

and, since $-F(\hat{x},\bar{x}) \leq \varepsilon$, we have

$$d(\hat{x},\bar{x}) \leq n\varepsilon. \tag{9.31}$$

Theorem 9.10. *Let (X,d) be a complete metric space and K be a nonempty compact subset of X. Let $F : K \times K \to \mathbb{R} \cup \{+\infty\}$ be a bifunction such that it is upper semicontinuous in the first argument and lower semicontinuous in the second argument. Assume that the following conditions hold:*

(i) $F(x,x) = 0$ for all $x \in K$;
(ii) $F(x,y) \leq F(x,z) + F(z,y)$ for all $x,y,z \in K$.

Then EP has a solution.

Proof. For each $n \in \mathbb{N}$, let $x_n \in K$ be an $\frac{1}{n}$-solution of EP (such point exists by Remark 9.10), that is,

$$F(x_n,y) \geq -\frac{1}{n}d(x_n,y) \quad \text{for all } y \in K.$$

Since K is compact, we can choose a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ such that $x_{n_k} \to \bar{x}$ as $n \to \infty$. Then by the upper semicontinuity of $F(\cdot,y)$, we have

$$F(\bar{x},y) \geq \limsup_{n\to\infty}\left(F(x_{n_k},y)+\frac{1}{n_k}d(x_{n_k},y)\right) \quad \text{for all } y \in K,$$

thereby proving that $\bar{x}$ is a solution of EP.

Example 9.3. The bifunction $F : [0,1]\times[0,1] \to \mathbb{R}$ defined as $F(x,y) = x^2 - y^2$ satisfies the assumption of Theorem 9.10.

We mention some equivalences among extended Ekelend's variational principle, extended Takahashi's minimization theorem, Caristi-Kirk fixed point theorem for set-valued maps, and Oettli-Théra theorem.

Theorem 9.11. *Let (X,d) be a complete metric space. Let $F : X \times X \to \mathbb{R}\cup\{+\infty\}$ be a bifunction such that it is lower semicontinuous in the second argument and the following conditions hold:*

(i) $F(x,x) = 0$ for all $x \in X$;
(ii) $F(x,y) \leq F(x,z)+F(z,y)$ for all $x,y,z \in X$.

Assume that there exists $\hat{x} \in X$ such that $\inf_{x\in X} F(\hat{x},x) > -\infty$. Let

$$\hat{S} := \{x \in X : F(\hat{x},x)+d(\hat{x},x) \leq 0\} \tag{9.32}$$

(From (i) it follows that $\hat{x} \in \hat{S} \neq \emptyset$). Then the following statements are equivalent:

(a) **(Extended Ekeland's Variational Principle).** *There exists $\bar{x} \in \hat{S}$ such that*

$$F(\bar{x},x)+d(\bar{x},x) > 0 \quad \text{for all } x \neq \bar{x}. \tag{9.33}$$

(b) **(Extended Takahashi's Minimization Theorem).** *Assume that*

$$\begin{cases} \text{for every } \hat{x} \in \hat{S} \text{ with } \inf_{x\in X} F(\hat{x},x) < 0 \text{ there exists} \\ x \in X \text{ such that } F(\hat{x},x)+d(\hat{x},x) \leq 0 \quad \text{for all } x \neq \hat{x}. \end{cases} \tag{9.34}$$

Then there exists $\bar{x} \in \hat{S}$ such that $F(\bar{x},x) \geq 0$ for all $x \in X$.

(c) **(Caristi–Kirk Fixed Point Theorem).** *Let $T : X \rightrightarrows X$ be a set-valued map such that*

$$\begin{cases} \text{for every } \hat{x} \in \hat{S} \text{ there exists} \\ x \in T(\hat{x}) \text{ satisfying } F(\hat{x},x)+d(\hat{x},x) \leq 0. \end{cases} \tag{9.35}$$

Then there exists $\bar{x} \in \hat{S}$ such that $\bar{x} \in T(\bar{x})$.

(d) **(Oettli–Théra Theorem).** *Let $D \subset X$ have the property that*

$$\begin{cases} \text{for every } \hat{x} \in \hat{S}\setminus D \text{ there exists} \\ x \in X \text{ such that } F(\hat{x},x)+d(\hat{x},x) \leq 0 \quad \text{for all } x \neq \hat{x}. \end{cases} \tag{9.36}$$

Then there exists $\bar{x} \in \hat{S}\cap D$.

Proof. (a) $\Rightarrow$ (d): Let (a) and the hypothesis of (d) hold. Then (a) gives $\bar{x} \in \hat{S}$ such that $F(\bar{x},x) + d(\bar{x},x) > 0$ for all $x \neq \bar{x}$. From (9.36), we have $\bar{x} \in D$. Hence $\bar{x} \in \hat{S} \cap D$ and (d) holds.

(d) $\Rightarrow$ (a): Let (d) hold. For all $\hat{x} \in X$, define

$$\Gamma(\hat{x}) = \{x \in X : F(\hat{x},x) + d(\hat{x},x) \leq 0,\ x \neq \hat{x}\}.$$

Choose $D := \{\hat{x} \in X : \Gamma(\hat{x}) = \emptyset\}$. If $\hat{x} \notin D$, then from the definition of D, there exists $x \in \Gamma(\hat{x})$. Hence (9.36) is satisfied, and by (d), there exists $\bar{x} \in \hat{S} \cap D$. Then $\Gamma(\bar{x}) = \emptyset$, that is, $F(\hat{x},x) + d(\hat{x},x) > 0$ for all $x \neq \bar{x}$. Hence (a) holds.

(b) $\Rightarrow$ (d): Suppose that both (b) and the hypothesis of (d) hold. Assume contrary that $\hat{x} \notin D$ for all $\hat{x} \in \hat{S}$. Then by (9.36) for all $\hat{x} \in \hat{S}$

$$\text{there exists } x \neq \hat{x} \text{ with } F(\hat{x},x) + d(\hat{x},x) \leq 0. \tag{9.37}$$

Hence (9.35) is satisfied. By (b), there exists $\bar{x} \in \hat{S}$ such that $F(\bar{x},x) \geq 0$ for all $x \in X$. This implies that $F(\bar{x},x) + d(\bar{x},x) > 0$ for all $x \in X$, $x \neq \bar{x}$, a contradiction with (9.37). Hence $\hat{x} \in D$ for some $\hat{x} \in \hat{S}$ and (d) holds.

(d) $\Rightarrow$ (b): Suppose that both (d) and the hypothesis of (b) hold. Choose $D := \{\hat{x} \in X : \inf_{x \in X} F(\hat{x},x) \geq 0\}$. Then (9.36) follows from (9.34), and (d) furnishes some $\bar{x} \in \hat{S} \cap D$. It follows from the definition of D that $\inf_{x \in X} F(\bar{x},x) \geq 0$. Hence (b) holds.

(c) $\Rightarrow$ (d): Let (c) and the hypothesis of (d) hold. Define a set-valued map $T : X \rightrightarrows X$ by

$$T(\hat{x}) = \{x \in X : x \neq \hat{x}\}.$$

Assume contrary that $\hat{x} \notin D$ for all $\hat{x} \in \hat{S}$. Then (9.35) follows from (9.36), and by (c) there exists $\bar{x} \in T(\bar{x})$. But this is clearly impossible from the definition of T. Hence $\hat{x} \in D$ for some $\hat{x} \in \hat{S}$ and (d) holds.

(d) $\Rightarrow$ (c): Suppose that both (d) and the hypothesis of (c) hold. Choose $D := \{\hat{x} \in X : \hat{x} \in T(\hat{x})\}$. Then (9.36) follows from (9.35), and (d) furnishes some $\bar{x} \in \hat{S} \cap D$ which, from the definition of D, necessarily belongs to $T(\bar{x})$. Hence (c) holds.

Remark 9.11.

(i) Theorem 9.11 (a) is nothing but Theorem 9.9.
(ii) If we choose $F(x,y) = F(y) - f(x)$, where f is lower semicontinuous functional on X, then the conditions (i) and (ii) of Theorem 9.9 are satisfied. In this case, Theorem 9.9 and Theorem 9.11 (b) reduce to Theorem 9.2 and Theorem 9.8, respectively.
(iii) Theorem 9.11 (b) guarantees the existence of a solution of EP without any compactness or convexity assumption on the underlying set.

The following condition was introduced by Oettli and Théra [69].

Definition 9.4. We say that $\hat{x} \in X$ satisfies (OT) condition if and only if every sequence $\{x_n\} \subset X$ satisfying $F(\hat{x},x_n) \leq 1/n$ for all $n \in \mathbb{N}$ and $F(x_n,x) + (1/n)d(x_n,x) \geq 0$ for all $x \in X$ and $n \in \mathbb{N}$, has a convergent subsequence.

Theorem 9.12. *Let (X,d) be a complete metric space. Let $F : X \times X \to \mathbb{R}$ be upper semicontinuous in the first argument and lower semicontinuous in the second argument. Assume that the following conditions hold:*

(i) $F(x,x) = 0$ *for all* $x \in X$;
(ii) $F(x,y) \leq F(x,z) + F(z,y)$ *for all* $x,y,z \in X$.

Then the following statements hold.

(a) If for some $\hat{x} \in X$, $\inf_{x \in X} F(\hat{x},x) > -\infty$ *and* $\hat{x}$ *satisfies (OT) condition, then there exists* $\bar{x} \in X$ *such that* $f(\bar{x},y) \geq 0$ *for all* $y \in X$.
(b) If for some $\hat{x} \in X$, $F(\hat{x},x) \geq 0$ *for all* $x \in X$, *and* $\hat{x}$ *satisfies (OT) condition, then for every* $\alpha > 0$: *either* $\inf\{F(\hat{x},x) : x \in X,\ d(\hat{x},x) = \alpha\} > 0$, *or there exists* $x_\alpha \in X$ *with* $F(\hat{x},x_\alpha) = \alpha$ *and* $F(\hat{x},x_\alpha) = 0$ *(implying* $F(x_\alpha,x) \geq 0$ *for all* $x \in X$*).*

Proof. (a) From Remark 9.10 (ii), for all $n \in \mathbb{N}$ there exists $x_n \in X$ such that, correspondingly with (9.29), (9.30),

$$F(x_n,x) + \frac{1}{n} d(x_n,x) \geq 0 \quad \text{for all } x \in X, \tag{9.38}$$

$$F(\hat{x},x_n) \leq 0.$$

By (OT) condition, there exists a subsequence of $\{x_n\}$ converging towards some $\hat{x} \in X$. Then from (9.38) and the upper semicontinuity of $F(\cdot,x)$, we have

$$F(\bar{x},x) \geq 0 \quad \text{for all } x \in X.$$

For the proof of (b), we refer to [69].

This page intentionally left blank

Appendix A
Some Basic Inequalities

Proposition A.1. *For any $x \geq 0$, the function $f(x) = \frac{x}{1+x}$ is monotonically increasing.*

Proof. Let $y > x \geq 0$. Then

$$\frac{1}{1+y} < \frac{1}{1+x} \text{ and so } 1 - \frac{1}{1+y} > 1 - \frac{1}{1+x},$$

that is,

$$\frac{y}{1+y} > \frac{x}{1+x}.$$

Theorem A.1. *For any two real numbers x and y, the following inequality holds:*

$$\frac{|x+y|}{1+|x+y|} \leq \frac{|x|}{1+|x|} + \frac{|y|}{1+|y|}. \tag{A.1}$$

Proof. Let x and y have the same sign. Without loss of generality, we may assume that $x \geq 0$ and $y \geq 0$. Then,

$$\begin{aligned} \frac{|x+y|}{1+|x+y|} &= \frac{x+y}{1+x+y} = \frac{x}{1+x+y} + \frac{y}{1+x+y} \\ &\leq \frac{x}{1+x} + \frac{y}{1+y} = \frac{|x|}{1+|x|} + \frac{|y|}{1+|y|}. \end{aligned}$$

If x and y have different signs, then we may assume that $|x| > |y|$. Then $|x+y| \leq |x|$. From Proposition A.1, we have

$$\frac{|x+y|}{1+|x+y|} \leq \frac{|x|}{1+|x|} \leq \frac{|x|}{1+|x|} + \frac{|y|}{1+|y|}.$$

Theorem A.2 (Hölder's Inequality). *Let $x_i \geq 0$ and $y_i \geq 0$ for all $i = 1, 2, \ldots, n$ and let $p > 1$ and $q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then*

$$\sum_{i=1}^{n} x_i y_i \leq \left(\sum_{i=1}^{n} x_i^p \right)^{1/p} \left(\sum_{i=1}^{n} y_i^q \right)^{1/q}. \tag{A.2}$$

When $p = q = 2$, then the above inequality becomes

$$\sum_{i=1}^{n} x_i y_i \leq \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2} \left(\sum_{i=1}^{n} y_i^2 \right)^{1/2}. \tag{A.3}$$

This inequality is known as Cauchy-Schwarz inequality.

Theorem A.3 (Minkowski's Inequality). *Let $x_i \geq 0$ and $y_i \geq 0$ for all $i = 1, 2, \ldots, n$ and let $p > 1$. Then*

$$\left(\sum_{i=1}^{n} (x_i + y_i)^p\right)^{1/p} \leq \left(\sum_{i=1}^{n} x_i^p\right)^{1/p} + \left(\sum_{i=1}^{n} y_i^p\right)^{1/p}. \tag{A.4}$$

Proof. If $p = 1$, then there is nothing to prove. So, we assume that $p > 1$. Note that

$$\sum_{i=1}^{n} (x_i + y_i)^p = \sum_{i=1}^{n} x_i (x_i + y_i)^{p-1} + \sum_{i=1}^{n} y_i (x_i + y_i)^{p-1}. \tag{A.5}$$

Let $q > 1$ be such that $\dfrac{1}{p} + \dfrac{1}{q} = 1$. From equation (A.5) and Hölder's inequality, we obtain

$$\begin{aligned}
\sum_{i=1}^{n} (x_i + y_i)^p &\leq \left(\sum_{i=1}^{n} x_i^p\right)^{1/p} \left(\sum_{i=1}^{n} (x_i + y_i)^{(p-1)q}\right)^{1/q} \\
&\quad + \left(\sum_{i=1}^{n} y_i^p\right)^{1/p} \left(\sum_{i=1}^{n} (x_i + y_i)^{(p-1)q}\right)^{1/q} \\
&= \left\{\left(\sum_{i=1}^{n} x_i^p\right)^{1/p} + \left(\sum_{i=1}^{n} y_i^p\right)^{1/p}\right\} \left(\sum_{i=1}^{n} (x_i + y_i)^p\right)^{1/q}.
\end{aligned}$$

If $\sum_{i=1}^{n} (x_i + y_i)^p = 0$, then obviously (A.4) holds. If $\sum_{i=1}^{n} (x_i + y_i)^p \neq 0$, then by dividing the above inequality by $\left(\sum_{i=1}^{n} (x_i + y_i)^p\right)^{1/q}$, we obtain (A.4).

Theorem A.4 (Minkowski's Inequality for Infinite Sums). *Let $p > 1$ and let $\{x_n\}$ and $\{y_n\}$ be sequences of nonnegative terms such that $\sum_{n=1}^{\infty} x_n^p$ and $\sum_{n=1}^{\infty} y_n^p$ are convergent. Then $\sum_{n=1}^{\infty} (x_n + y_n)^p$ is convergent. Moreover,*

$$\left(\sum_{n=1}^{\infty} (x_n + y_n)^p\right)^{1/p} \leq \left(\sum_{n=1}^{\infty} x_n^p\right)^{1/p} + \left(\sum_{n=1}^{\infty} y_n^p\right)^{1/p}. \tag{A.6}$$

Proof. For any positive integer m, from Theorem A.3, we have

$$\begin{aligned}
\left(\sum_{n=1}^{m} (x_n + y_n)^p\right)^{1/p} &\leq \left(\sum_{n=1}^{m} x_n^p\right)^{1/p} + \left(\sum_{n=1}^{m} y_n^p\right)^{1/p} \\
&\leq \left(\sum_{n=1}^{\infty} x_n^p\right)^{1/p} + \left(\sum_{n=1}^{\infty} y_n^p\right)^{1/p}.
\end{aligned}$$

Since $\left\{\left(\sum_{n=1}^{m}(x_n+y_n)^p\right)^{1/p}\right\}$ is an increasing sequence of nonnegative real numbers, it is bounded above by the sum

$$\left(\sum_{n=1}^{\infty}x_n^p\right)^{1/p}+\left(\sum_{n=1}^{\infty}y_n^p\right)^{1/p}.$$

It follows that $\sum_{n=1}^{\infty}(x_n+y_n)^p$ is convergent and that the inequality (A.6) holds.

Theorem A.5. *Let $p>1$. For any $a\geq 0$ and $b\geq 0$, we have*

$$(a+b)^p\leq 2^{p-1}(a^p+b^p). \tag{A.7}$$

Proof. If either a or b is 0, then nothing to prove. Suppose that $a>0$ and $b>0$. Since the function $x\mapsto x^p$ defined on the set of all positive numbers is convex when $p>1$, we have

$$\left(\frac{a+b}{2}\right)^p\leq\frac{a^p+b^p}{2}\quad\text{equivalently}\quad(a+b)^p\leq 2^{p-1}(a^p+b^p).$$

Theorem A.6 (Weierstrass's Theorem). *Let $f:[a,b]\to\mathbb{R}$ be a continuous function. Then there exists a sequence $\{p_n\}$ of polynomials with real coefficients that converges uniformly to f on $[a,b]$, that is, for all $\varepsilon>0$, there exists a positive integer N such that for all $t\in[a,b]$*

$$|p_n(t)-f(t)|<\varepsilon\quad\textit{whenever}\quad n>N.$$

This page intentionally left blank

Appendix B
Partial Ordering

Definition B.1 (Partial Ordering). A binary relation $\preccurlyeq$ on a nonempty set X is called *partial ordering* if the following conditions hold:

(i) For all $a \in X$, $a \preccurlyeq a$; (Reflexivity)
(ii) If $a \preccurlyeq b$ and $b \preccurlyeq a$, then $a = b$; (Antisymmetry)
(iii) If $a \preccurlyeq b$ and $b \preccurlyeq c$, then $a \preccurlyeq c$; (Transitivity)

Definition B.2 (Partially Ordered Set). A nonempty set X is called *partially ordered* if there is a partial ordering on X.

Remark B.1. The word 'partially' emphasizes that X may contain elements a and b for which neither $a \preccurlyeq b$ nor $b \preccurlyeq a$ holds. In this case, a and b are called *incomparable elements*. If $a \preccurlyeq b$ or $b \preccurlyeq a$ (or both), then a and b are called *comparable elements*.

Definition B.3 (Totally Ordered Set). A partially ordered set X is said to be *totally ordered* if every two elements of X are comparable.

In other word, partially ordered set X is totally ordered if it has no incomparable elements.

Definition B.4. Let M be a nonempty subset of a partially ordered set X.

- The element $x \in X$ is called an *upper bound* of M if $a \preccurlyeq x$ for all $a \in M$.
- The element $x \in X$ is called a *lower bound* of M if $x \preccurlyeq a$ for all $a \in M$.
- The element $x \in M$ is called a *maximal element* of M if $x \preccurlyeq a$ implies $x = a$.

Remark B.2. A subset of a partially ordered set M may or may not have an upper or lower bound. Also, M may or may not have maximal elements. Note that a maximal element need not be an upper bound.

Example B.1. (a) Let X be the set of all real numbers and let $a \leq b$ have its usual meaning. Then X is totally ordered and it has no maximal element.

(b) Let $X = \mathbb{N}$, the set of all natural numbers. Let $m \preccurlyeq n$ mean that m divides n. Then $\preccurlyeq$ is a partial ordering on X.

(c) Let $\mathscr{P}(X)$ be the power set, that is, set of all subsets of a set X. Let $A \preccurlyeq B$ mean $A \subset B$. Then $\mathscr{P}(X)$ is a partially ordered set and the only maximal element of $\mathscr{P}(X)$ is X.

(d) Let X be the set of all ordered n-tuples $x = (x_1, x_2, \ldots x_n)$, $y = (y_1, y_2, \ldots y_n)$, $\ldots$ of real numbers and let $x \preccurlyeq y$ mean $x_i \leq y_i$ for all $i = 1, 2, \ldots, n$. Then $\preccurlyeq$ is a partial ordering on X.

Lemma B.1 (Zorn's Lemma). *Let X be a nonempty partially ordered set in which every totally ordered set has an upper bound. Then X has at least one maximal element.*

This page intentionally left blank

Appendix C
Nested Interval Property

Least Upper Bound Axiom. If A is a set of real numbers bounded from above, then A has a least upper bound.

Theorem C.1 (Nested Interval Property). *Let $I_1 = [a_1, b_1]$, $I_2 = [a_2, b_2]$, ... be a sequence of nested closed and bounded intervals, that is, $I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots \supseteq I_n \supseteq I_{n+1} \supseteq \cdots$. Then there exists at least one point common to every interval, that is, $\bigcap_{i=1}^{\infty} I_i \neq \emptyset$.*

Proof. Since $I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots \supseteq I_n \supseteq I_{n+1} \supseteq \cdots$, we have

$$a_1 \leq a_2 \leq \cdots \quad \text{and} \quad \cdots \leq b_2 \leq b_1.$$

We claim that $a_m < b_n$ for every $m, n \in \mathbb{N}$

If $m > n$ then $a_m < b_m \leq b_n$, and if $m \leq n$ then $a_m \leq a_n < b_n$. Thus each b_n is an upper bound for the set $A = \{a_1, a_2, a_3, \ldots\}$ of left end points. By the Least Upper Bound Axiom of $\mathbb{R}$, A has a least upper bound, that is, $\sup(A)$ exists, say, $p = \sup(A)$.

Since each b_n is an upper bound for A and p is the least upper bound, we have $p \leq b_n$ for each $n \in \mathbb{N}$. Furthermore, since p is an upper bound for $A = \{a_1, a_2, \ldots\}$, we have $a_n \leq p$ for each $n \in \mathbb{N}$. But

$$a_n \leq p \leq b_n \quad \text{implies} \quad p \in I_n = [a_n, b_n].$$

Hence p is common to every interval.

Remark C.1. It is necessary that the intervals in the above theorem be closed and bounded, otherwise the theorem is not true.

Example C.1. Let $A_1 = (0, 1]$, $A_2 = (0, \frac{1}{2}]$, ... , $A_n = (0, \frac{1}{n}]$, ..., then $\{A_n\}$ is the sequence of open-closed intervals. It is clear that the sequence $\{A_n\}$ is nested, that is, $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$. But the intersection of the intervals is empty, that is, $\bigcap_{n=1}^{\infty} A_n = \emptyset$.

Example C.2. Let $\{A_n\}$ be the following sequence of closed infinite intervals

$$A_1 = [1, \infty),\ A_2 = [2, \infty),\ \ldots,\ A_n = [n, \infty), \ldots.$$

Then clearly $A_1 \supset A_2 \supset A_3 \supset \cdots$, that is, the sequence of intervals is nested. But $\bigcap_{n=1}^{\infty} A_n = \emptyset$.

Theorem C.2 (Heine-Borel Theorem). *Every closed and bounded interval of the real line $\mathbb{R}$ is compact.*

Proof. Let $I_1 = [a_1, b_1]$ be a closed and bounded subset of $\mathbb{R}$. Assume that I_1 is not compact. Then there exists an open cover $\mathscr{F} = \{G_\alpha\}_{\alpha \in \Lambda}$ of I_1 which does have

a finite subcover. Divide $I_1 = [a_1, b_1]$ into two closed subintervals of equal length $\left[a_1, \frac{a_1+b_1}{2}\right]$ and $\left[\frac{a_1+b_1}{2}, b_1\right]$.

Then, by assumption, at least one of these two intervals will not be covered by any finite subcover of the cover $\mathscr{F}$. Let us denote such interval by I_2 and write $I_2 = [a_2, b_2]$. Then $[a_2, b_2]$ is either $\left[a_1, \frac{a_1+b_1}{2}\right]$ or $\left[\frac{a_1+b_1}{2}, b_1\right]$.

Divide I_2 into two closed subintervals of equal length $\left[a_2, \frac{a_2+b_2}{2}\right]$ and $\left[\frac{a_2+b_2}{2}, b_2\right]$. Again, by assumption, at least one of these two intervals will not be covered by any finite subcover of the cover $\mathscr{F}$. Let us denote such interval by I_3 and write $I_3 = [a_3, b_3]$.

Continuing in this way, we obtain a sequence of closed intervals having the following properties:

(i) $I_n \supset I_{n+1}$ for all $n \in \mathbb{N}$,
(ii) I_n is closed for all $n \in \mathbb{N}$,
(iii) I_n is not covered by any finite subclass of $\mathscr{F}$,
(iv) $\delta(I_n) \to 0$ as $n \to \infty$, where $\delta(I_n)$ denotes the length of the interval I_n.

Hence by the nest closed interval property,

$$\bigcap_{n=1}^{\infty} I_n \neq \emptyset.$$

So there exists a number $x \in \bigcap_{n=1}^{\infty} I_n$ which implies that $x \in I_n$ for each $n \in \mathbb{N}$. In particular, $x \in I_1$.

Since $\mathscr{F}$ is an open cover of I_1, there exists $G_{\alpha_0} \in \mathscr{F}$ such that $x \in G_{\alpha_0}$. Since G_{α_0} is an open set, there exists an open neighborhood $(x-\varepsilon, x+\varepsilon)$ such that

$$x \in (x-\varepsilon, x+\varepsilon) \subseteq G_{\alpha_0}.$$

Now, $\delta(I_n) \to 0$ as $n \to \infty$, we can choose a positive integer n_0 large enough that $\delta(I_{n_0}) < \varepsilon$. This implies that

$$I_{n_0} \subseteq (x-\varepsilon, x+\varepsilon) \subseteq G_{\alpha_0}.$$

This contradicts to our assumption that no I_n is covered by a finite number of members of $\mathscr{F}$. Hence $[a_1, b_1]$ is compact.

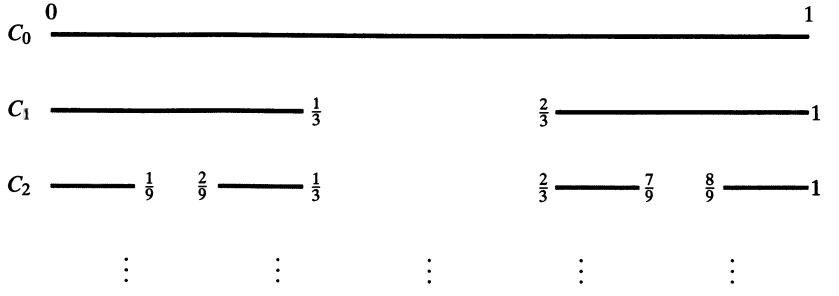
Definition C.1 (Cantor Set). Let C_0 denote the interval $[0,1]$. Remove from C_0 in succession:

(i) the open interval $\left(\frac{1}{3}, \frac{2}{3}\right)$, the middle third of the interval in C_0, leaving behind the set

$$C_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right].$$

(ii) the open intervals $\left(\frac{1}{9}, \frac{2}{9}\right)$ and $\left(\frac{7}{9}, \frac{8}{9}\right)$, the middle thirds of the two closed intervals $\left[0, \frac{1}{3}\right]$ and $\left[\frac{2}{3}, 1\right]$ in C_1, leaving behind the set

$$C_2 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right].$$



Continue in this way, we get a sequence $\{C_n\}$ of sets, where C_n is obtained from C_{n-1} by removing the (open) middle thirds of the $2^{(n-1)}$ disjoint closed intervals of which C_{n-1} is composed of, that is,

$$C_n = C_{n-1} - \bigcup_{k=0}^{\infty} \left(\frac{1+3k}{3^n}, \frac{2+3k}{3^n} \right).$$

The intersection

$$C = \bigcap_{n \in \mathbb{N}} C_n$$

is called the *Cantor set.*

Each of the sets C_n is nonempty, closed and bounded. Also, $C_n \subset C_{n-1}$ for each n. Hence the set C is nonempty, closed and bounded.

Exercise C.1. Prove that the Cantor set C is compact.

Proof. Since each C_n is the union of 2^n disjoint closed intervals, so that each C_n is closed. Also, arbitrary intersection of closed sets is closed, we have

$$C = \bigcap_{n=1}^{\infty} C_n \text{ is closed.}$$

Also, C is bounded, because $C \subset [0,1]$. Hence by Heine-Borel Theorem C.2, C is compact.

References

1. S. Al-Homidan, Q. H. Ansari and J.-C. Yao, Some generalizations of Ekeland-type variational principle with applications to equilibrium problems and fixed point theory, *Nonlinear Anal.*, **69** (2008), 126–139.
2. A. G. Aksoy and M. A. Khamsi, *Real Analysis Through Problem Solving*, 2008.
3. Q. H. Ansari, N. C. Wong and J. C. Yao, The existence of nonlinear inequalities, *Appl. Math. Lett.*, **12** (5) (1999), 89–92.
4. J.-P. Aubin *Mathematical Methods of Game and Economic Theory*, North-Holland Publishing Company, Amsterdam, New York, Oxford, 1979.
5. J.-P. Aubin and A. Cellina, *Differential Inclusions, Set-Valued Maps and Viability Theory*, Springer-Verlag, New York, Heidelberg, Berlin, 1984.
6. J.-P. Aubin and I. Ekeland, *Applied Nonlinear Analysis*, John Wiley & Sons, Inc., New York, Chichester, Brisbane, Toronto, Singapore, 1984.
7. J.-P. Aubin and H. Frankowska, *Set-Valued Analysis*, Birkhäuser, Boston, Basel, Berlin, 1990.
8. J. S. Bae and S. Park, Remarks on the Caristi-Kirk fixed point theorem, *Bull. Korean Math. Soc.*, **19** (1983), 57–60.
9. M. Barnsley, *Fractals Everywhere*, Academic Press, Inc., Harcourt Brace Jovanovich, Publishers, Boston, New York, London, Tokyo, Toronto, 1988.
10. M. Bianchi and S. Schaible, Generalized monotone bifunctions and equilibrium problems, *J. Optim. Theory Appl.*, **90** (1996), 31–43.
11. M. Bianchi and S. Schaible, Equilibrium problems under generalized convexity and generalized monotonicity, *J. Global Optim.*, **30** (2004), 121–134.
12. M. Bianchi, G. Kassay and R. Pini, Existence of equilibria via Ekeland's principle, *J. Math. Anal. Appl.*, **305** (2005), 502–512.
13. C. Berge, *Topological Spaces: Including a treatment of Multi-valued Functions, Vector Spaces and Convexsity*, Oliver & Boyd Ltd., Edinburgh and London, 1963.
14. E. Blum and W. Oettli, From optimization and variational inequalities to equilibrium problems, *Math. Student*, **63** (1-4) (1994), 123–145.
15. J. M. Borwein and Q. J. Zhu, *Techniques of Variational Analysis*, Springer-Verlag, Berlin, Heidelberg, New York, 2005.
16. C. Bosch, A. Garcia and C. L. Garcia, An extension of Ekeland's variational principle to locally complete spaces, *J. Math. Anal. Appl.*, **328** (2007), 106–108.
17. G. Bouligand, Sur la semi-continuité d'inclusions et quelques sujets connexes, *Enseignement Math.*, **31** (1932), 14–22.
18. D. Boyd and J. S. W. Wong, On nonlinear contractions, *Proc. Amer. Math. Soc.*, **20** (1969), 458–464.
19. H. Brézis, L. Nirenberg and G. Stampacchia, A remark on Ky Fan's minimax principle, *Boll. Un. Mat. Ital.*, **6** (1972), 293–300.
20. V. Bryant, *Metric Spaces, Iteration and Applications*, Cambridge University Press, Cambridge, New York, Sydney, 1985.
21. R. S. Burachik and A. N. Iusem, *Set-Valued Mappings and Enlàrgments of Monotone Operators*, Springer-Verlag, New York, 2008.
22. J. Caristi, Fixed point theorems for mappings satisfying inwardness conditions, *Trans. Amer. Math. Soc.*, **215** (1976), 241–251.
23. J. Caristi and W. A. Kirk, Geometric fixed point theory and inwardness conditions, in *The Geometry of Metric and Linear Spaces*, Lecture Notes in Mathematics, **490**, Springer-Verlag, Berlin, Heidelberg, New York, 1975, **pp.** 74–83.
24. O. Chadli, Z. Chbani and H. Riahi, Equilibrium problems with generalized monotone bifunctions and applications to variational inequalities, *J. Optim. Theory Appl.*, **105** (2000), 299–323.
25. O. Chadli, N. C. Wong and J. C. Yao, Equilibrium problems with applications to eigenvalue problems, *J. Optim. Theory Appl.*, **117** (2003), 245–266.
26. B. Choudhary and S. Nanda, *Functional Analysis with Applications*, Wiley Easyern Limited, New Delhi, Banagalore, Bombay, 1989.
27. E. Cinlar and R. J. Vanderbei, *Mathematical Methods of Engineering Analysis*, Lecture Notes, 2000.
28. F. Clarke, Pointwise contraction criteria for the existence of fixed points, *MRC Technical Report* **1658**, July 1976, University of Wisconsin, Madison, Wisconsin, 1976.

29. E. T. Copson, *Metric Spaces*, Cambridge University Press, Cambride, U.K., 1968.
30. S. Dancs, M. Hegedűs and P. Medvegyev, A general ordering and fixed-point principle in complete metric space, *Acta Sci. Math.*, **46** (1983), 381–388.
31. J. Daneš, A geometric theorem useful in nonlinear analysis, *Boll. Un. Mat. Ital.*, **6** (1972), 369–372.
32. J. Daneš, Equivalence of some geometric and related results of nonlinear functional analysis, *Comment. Math. Univ. Carolinae*, **26** (1985), 443–454.
33. I. Ekeland, Sur les prolèms variationnels, *C. R. Acad. des Sci. Paris*, **275** (1972), 1057–1059.
34. I. Ekeland, On the variational principle, *J. Math. Anal. Appl.*, **47** (1974), 324–353.
35. I. Ekeland, On convex minimization problems, *Bull. Amer. Math. Soc.*, **1**(3) (1979), 445–474.
36. F. Facchinei and J.-S. Pang, *Finite Dimensional Variational Inequalities and Complementarity Problems*, I, Springer-Verlag, New York, Berlin, Heidelberg, 2003.
37. K. Fan, Extensions of two point theorems of F.E. Browder, *Math. Z.*, **112** (1969), 234–240.
38. K. Fan, A minimax inequlity and applications, in *Inequality III*, edited by O. Shisha, Academic Press, New York, (1972), **pp.** 103–113.
39. D. G. De Figueiredo, *The Ekeland Variational Principle with Applications and Detours*, Tata Institute of Fundamental Research, Bobmay, 1989.
40. F. Flores-Bazán, Existence theorems for generalized noncoercive equilibrium problems: The quasi-convex case, *SIAM J. Optim.*, **11** (2000), 675–690.
41. F. Flores-Bazán, Existence theory for finite-dimensional pseudomonotone equilibrium problems, *Acta Appl. Math.*, **77** (2003), 249–297.
42. P.G. Georgiev, The strong Ekeland variational principle, the strong drop theorem and applications, *J. Math. Anal. Appl.*, **131** (1988), 1–21.
43. J. R. Giles, *Convex Analysis with Application in the Differention of Convex Functions*, Research Notes in Mathematics, **58**, Pitman Advanced Publishing Program, Boston, London, Melbourne, 1982.
44. J. R. Giles, *Introduction to the Analysis of Metric Spaces*, Cambridge University Press, Cambridge, New York, Rochell, Melbourne, Sydney, 1987.
45. A. Göfert, Chr. Tammer, H. Riahi and C. Zălinescu, *Variational Methods in Partially Ordered Spaces*, Springer-Verlag, New York, Berlin, Heidelberg, 2003.
46. L. Górniewicz, *Topological Fixed Point Theory of Multivalued Mappings*, Kluwer Academic Publishers, Dordrecht, Boston, London, 1999.
47. N. Hadjisavvas, S. Komlósi and S. Schaible, *Handbook of Generalized Convexity and Generalized Monotonicity*, Springer-Verlag, Berlin, Heidelberg, New York, 2005.
48. A. Hamel, Remarks to an equivalent formulation of Ekeland's variational principle, *Optimization*, **31** (3) (1994), 233–238.
49. A. Hamel, Phelps' lemma, Daneš' drop theorem and Ekeland principle in locally convex spaces, *Proc. Amer. Math. Soc.*, **131** (2003), 3025–3038.
50. A. Hamel, Equivalents to Ekeland's variational principle in uniform spaces, *Nonlinear Anal.*, **62**(5) (2005), 913–924.
51. V. I. Istrătescu, *Fixed Point Theory: An Introduction*, D. Reidel Publicating Company, Dordrecht, Boston, London, 1981.
52. P. K. Jain and K. Ahmad, *Metric Spaces*, Narosa Publishing House Pvt. Ltd., New Delhi, 1993.
53. M. C. Joshi and R. K. Bose, *Some Topics in Nonlinear Functional Analysis*, Wiley Eastern Limited, New Delhi, Bangalore, Bombay, Calcutta, Madras, Hydrabad, 1985.
54. O. Kada, T. Suzuki and W. Takahashi, Nonconvex minimization theorems and fixed point theorems in complete metric spaces, *Math. Japonica*, **44**(2) (1996), 381–391.
55. S. Kakutani, A generalization of Brouwer's fixed point theorem, *Duke Math. J.*, **8** (1941), 457–459.
56. M. A. Khamsi and W. A. Kirk, *An Introdunction to Metric Spaces and Fixed Point Theory*, John Wiley & Sons, Inc., New York, Chichester, Brisbane, Singapore, Toronto, 2001.
57. E. Klein and A. C. Thompson, *Theory of Correspondences, Including Applications to Mathematical Economics*, John Wiley & Sons, Inc., New York, Chichester, Brisbane, Toronto, Singapore, 1984.
58. E. Kreyszig, *Introductory Functional Analysis with Applications*, John Wiley & Sons, New York, Chichester, Brisbane, Toronto, Singapore, 1978.
59. K. Kuratowski, Les fonctions semi-continues dans l'espace des ensembles fermés, *Fund. Math.*, **18** (1932), 148–159.

60. S. Lipschutz, *Theory and Problems of General Topology*, Schaum's Outlines Series, McGraw-Hill Book Company, New York, Toronto, Sydney, 1965.
61. V. L. Makarov and A. M. Rubinov, *Mathematical Theory of Economic, Dynamics and Equilibria*, Springer-Verlag, New York, Heidelberg, Berlin, 1977.
62. J. Matkowski, Nonlinear contractions in metrically convex spaces, *Publ. Math. Debrecen*, **45** (1994), 103–114.
63. L. McLinden, An application of Ekeland's theorem to minimax problems, *Nonlinear Anal. Theory, Meth. Appl.*, **6** (1982), 189–196.
64. S. B. Nadler, Jr., Multivalued contraction mappings, *Pacific J. Math.*, **30** (1969), 475–488.
65. J. F. Nash, Equilibrium points in n-person games, *Proc. Nat. Acad. Sci. USA*, **36** (1950), 48–49.
66. J. F. Nash, Noncooperative games, *Ann. Math.*, **54** (1951), 286–295.
67. A. B. Nemeth, A nonconvex vector minimization problem, *Nonlinear Anal. Theory, Meth. Appl.*, **10** (1986), 669–678.
68. H. Nikaido and K. Isoda, Note on noncooperative convex games, *Pafic J. Math.*, **5** (1955), 807–815.
69. W. Oettli and M. Théra, Equivalents of Ekeland's principle, *Bull. Austral. Math. Soc.*, **48** (1993), 385–392.
70. S. Park, Equivalent formulations of Ekeland's variational principle for approximate soltuions of minimization problems and their applications, in *Operator Equations and Fixed Point Theorems*, edited by S.P. Singh, V.M. Sehgal and J.H.W. Burry, MSRI-Korea Publishers 1, Soul, 1986, **pp.** 55–68.
71. S. Park, On generalizations of the Ekeland-type variational principles, *Nonlinear Anal.*, **39** (2000), 881–889.
72. J.-P. Penot, The drop theorem, the petal theorem and Ekeland's variational principle, *Nonlinear Anal. Theory, Meth. Appl.*, **10** (1986), 813–822.
73. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
74. R.T. Rockafellor, Directionally Lipschitz functions and subdifferential calculus, *Proc. London Math. Soc.*, **39** (1979), 331–355.
75. M. Ó. Searcóid, *Metric Spaces*, Springer-Verlag, London, 2007.
76. S. Shirali and H. Vasudeva, *Metric Spaces*, Springer-Verlag, London, 2006.
77. A. H. Siddiqi, *Applied Functional Analysis: Numerical Methods, Wavelete Methods, and Image Processing*, Marcel Dekker, Inc., New York, Basel, 2004.
78. G. F. Simmons, *Introduction to Topology and Modern Analysis*, McGraw-Hill, 1963.
79. S. Singh, B. Watson and P. Srivastava, *Fixed Point Theory and Best Approximation: The KKM-Map Principle*, Kluwer Avademic Publishers, Dordrecht, Boston, London, 1997.
80. R. E. Smithson, Multifunctions, *Niew Archief Voor-Wiskunde*, **(3)XX**, (1972), 31–53.
81. G. Stampacchia, Formes bilinéaires coercitives sur les ensembles convexes, *C. R. Acad. Sci. Paris*, **258** (1964), 4413–4416.
82. E. S. Suhubi, *Functional Analysis*, Kluwer Academic Publishers, Dordrecht, Boston, London, 2003.
83. F. Sullivan, A characterization of complete metric spaces, *Proc. Amer. Math. Soc.*, **83** (1981), 345–346.
84. T. Suzuki, Generalized distnace and existence theorems in complete metric spaces, *J. Math. Anal. Appl.*, **253** (2001), 440–458.
85. T. Suzuki, The strong Ekeland variational principle, *J. Math. Anal. Appl.*, **320** (2006), 787–794.
86. W. Takahashi, Existence theorems generalizing fixed point theorems for multivalued mappings, in *Fixed Point Theory and Applications*, edited by J.-B. Baillon and M. Théra, Pitman Research Notes in Mathematics, **252**, Longman, Harlow, 1991, **pp.** 397–406.
87. W. Takahashi, *Nonlinear Functional Analysis*, Yokohama Publishers, Yokohama, 2000.
88. Chr. Tammer, A generalization of Ekeland's variational principle, *Optimization*, **25** (1992), 129–141.
89. G.X.-Z. Yuan, *KKM Theory and Applications in Nonlinear Analysis*, Marcel Dekker, Inc., New York, Basel, 1999.
90. C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific, New Jersey, London, Singapore, Hong Kong, 2002.

Index

ε-net, 77
ε-solution, 178

Baire's category theorem, 64
Banach contraction theorem, 122
base, 55
  countable, 55
basis, 55
  countable, 55
Bolzano-Weierstrass property, 74
Bolzano-Weierstrass theorem, 74
boundary point, 27
bounded function, 132
bounded set, 13

Cantor's intersection theorem, 51
Caristi and Kirk fixed point theorem, 160
Cauchy sequence, 37
Cauchy-Schwarz inequality, 183
closed set, 24
closed sphere, 15
closure, 24
compact space, 69
complete metric space, 42
completion, 53
component, 118
connected set, 111
connected space, 111
continuous function, 87
  uniformly, 94
contraction mapping, 122
contractive map, 122
convergent sequence, 33
cover, 67
  open, 67

dense set, 57
  everywhere, 57
derived set, 23
DHM theorem, 159
diameter of a set, 13
directional contraction, 170
disconnected space, 111
discrete metric space, 3
discrete two element space, 116
distance between a point and a set, 13
distance between sets, 13

equilibrium problem, 174
equivalent metric spaces, 98
equivalent metrics, 98

finite intersection property, 71
finite subcover, 67
first axiom space, 56
first category space, 64
first countable space, 56
fixed point, 121, 157

Hölder's inequality, 183
Hausdorff distance, 153
Heine-Borel theorem, 76
homeomorphism, 97

interior of a set, 19
interior point, 19
isolated point, 24
isometry, 54, 97

Lebesgue number, 80
Lebesgue's covering lemma, 80
limit point, 23, 33
Lipschitz constant, 122
Lipschtiz continuous function, 122
lower bound, 187
lower semicontinuous function, 136

maximal element, 187
metric, 1
  discrete, 3
  equivalent, 98
  Hausdorff, 153

space, 1
usual, 2
metric projection, 141
metric space, 1
complete, 42
equivalent, 98
metric subspace, 29
minimization problem, 174
Minkowski's inequality, 184
Minkowski's inequality for infinite sums, 184
multivalued map, 140

Nash equilibrium problem, 175
neighborhood of a point, 19
nonexpansive map, 122
nowhere dense set, 63

open neighborhood, 19
open set, 19
open sphere, 15

partial ordering, 187
partially ordered set, 187
Picard theorem, 129
proper function, 161
pseudometric, 9
pseudometric space, 9

reducible finite cover, 67
relative metric, 29

saddle point problem, 174
second axiom space, 56
second category space, 64
second countable space, 56
segment, 170
separable space, 60
separated sets, 109
sequence, 33
bounded, 35
Cauchy, 37
constant, 33
convergent, 33
divergent, 33
eventually constant, 33
sequentially compact set, 75
sequentially compact space, 75
set-valued map, 140
$\mathscr{H}$-continuous, 155
bounded, 151
closed, 149
continuous, 147
contraction, 157
generalized Picard-iteration, 159
hyperinjective, 141
hyperunivocal, 141
inverse, 143
Lipschtiz, 157
lower semicontinuous, 146
nonexpansive, 157
quasi-surjective, 141
upper semicontinuous, 146
subcover, 67
subsequence, 33

totally bounded space, 77
totally ordered set, 187

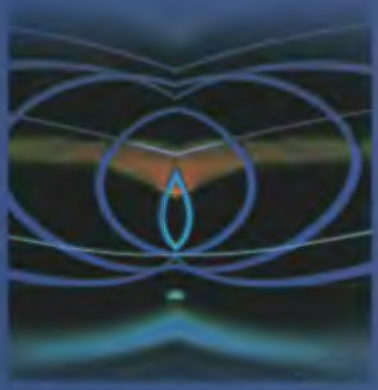
unbounded set, 13
uniformly continuous function, 94
upper bound, 187
upper semicontinuous function, 136

variational inequality problem, 175
Volterra equation, 131

Zorn's lemma, 187

METRIC SPACES

Including Fixed Point Theory and Set-valued Maps



Qamrul Hasan Ansari

METRIC SPACES is intended for undergraduate students offering a course of metric spaces and post graduate students offering a course of nonlinear analysis or fixed point theory. The first six chapters cover basic concepts of metric spaces, separable spaces, compact spaces, connected spaces and continuity of functions defined on a metric space. Chapter seven is devoted to the metric fixed point theory. Banach contraction theorem and several of its generalizations along with their applications and Caristi's fixed point theorem are also given in this chapter. The introductory set-valued analysis with special emphysis on continuity and fixed point theory of set-valued maps is given in chapter eight. One of the most useful and important results from nonlinear analysis is Ekeland's variational principle. This principle along with several of its equivalent forms, Takahashi's minimization theorem, introduction of theory of equilibrium problems and the equilibrium version of Ekeland's variational principle with several of its equivalent forms are presented in the last chapter.

This book will also be useful for researchers working in nonlinear analysis, optimization and theory of equilibrium problems.

Alpha Science International Ltd.
www.alphasci.com

ISBN-978-1-84265-655-6